

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

NWSETE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_dp protein - protein database search, using Smith-Waterman algorithm
Run on: Sun Sep 3 12:53:12 2000; MasPar time 29.26 Seconds
Tabular output not generated. 900.995 Million cell updates/sec

Title: >US-09-332-522B-8
Description: (1-1113) from US09332522B.pep
Perfect Score: 8012
Sequence: 1 MDITLMLNLIIDAPLDESMDLF.....LKECYOLMKSLGNGIGSVKA 1113

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23586105 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseqp

Statistics: Mean 40.405; Variance 208.262; scale 0.194

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1494	18.6	1147	1	Human SREBP-1a.	4.82e-108
2	1328	16.6	1141	1	Human SREBP-2.	2.09e-94
3	186	2.3	618	1	Human chromosome X TFE	6.04e-04
4	172	2.1	597	1	Human PRCC-TFE3 con	5.44e-03
5	172	2.1	700	1	Human PRCC-TFE3 fusion	5.44e-03
6	172	2.1	834	1	Human PRCC-TFE3 fusion	5.44e-03
7	153	2.0	383	1	Petunia Phb gene produ	2.19e-02
8	149	1.9	160	1	Max 11.	1.83e-01
9	149	1.9	2414	1	Transcription factor p	1.83e-01
10	149	1.9	2414	1	Cellular transcription	1.83e-01
11	148	1.8	151	1	Max 14.	2.13e-01
12	140	1.7	160	1	Breast cancer associat	1.45e+00
13	135	1.7	258	1	Max protein.	6.99e-01
14	134	1.7	1004	1	DEL protein.	1.68e+00
15	133	1.7	1004	1	A tumour suppressor ge	1.94e+00
16	133	1.7	1004	1	Human polyomelic 1 (5.32e+00
17	126	1.6	259	1	Elmeria cell surface	6.13e+00
18	125	1.6	297	1	Cotton fibrous tissue	3.46e+00
19	129	1.6	572	1	Mycobacterium tubercul	3.46e+00
20	129	1.6	763	1	Mycobacterium tubercul	3.46e+00
21	126	1.6	846	1	Human CLOCK protein.	5.32e+00
22	126	1.6	846	1	Human HSCLOCK polypept	5.32e+00
23	126	1.6	855	1	Mouse CLOCK protein.	5.32e+00

ALIGNMENTS

RESULT	ID	Accession	Standard	Protein	Length	Score	Alignment
AC	R66390	standard	Protein	1147 AA.	1	1.5	Human eps8.
DT	04-AUG-1995	(first entry)			1	1.5	Cotton fiber-specific
DE	Human SREBP-1a.				1	1.5	Apoptosis-blocking pro
KW	Probe: sterol regulatory element binding protein; regulatory protein;				1	1.5	Human Bcl-2 mutant pro
KW	cholesterol metabolism; sterol regulatory element-1; SREBP; SRP-1;				1	1.5	Streptococcus pneumonia
KW	SREBP-1; SREBP-2; basic-helix-loop-helix-leucine zipper; plasma;				1	1.5	Human chromosome 19 de
KW	transcription factor; low density lipoprotein; LDL; receptor;				1	1.5	Amino acids 71-464 of
KW	3-hydroxy-3-methylglutaryl coenzyme A synthase; cholesterol;				1	1.5	N-myc protein
KW	hypercholesterolaemia; enhancer.				1	1.5	Human basic helix-loop
OS	Homo sapiens.				1	1.5	Streptococcus pneumoni
EH	Key				1	1.5	Tobacco PABF protein.
FT	Region				1	1.5	Staphylococcus aureus
FT	Location/Qualifiers				1	1.5	Maize protein encoded
FT	1..60				1	1.5	Human TLE-2.
FT	/note= "Acidic region"				1	1.5	Human TLE-2.
FT	61..177				1	1.5	EBV gp350/220
FT	/note= "Pro/Ser rich domain"				1	1.5	Fibronectin-binding pr
FT	324..394				1	1.5	CREB binding protein.
FT	/label= bHLH-zip				1	1.5	Human mitotin amino ac
FT	324..333				1	1.5	Human mitotin.
FT	/note= "Basic region"				1	1.5	Kinetochores protein CE
FT	337..349				1	1.5	
FT	/note= "Helix 1"				1	1.5	
FT	358..373				1	1.5	
FT	/note= "Helix 2"				1	1.5	
FT	359..401				1	1.5	
FT	/note= "Putative Leucine zipper"				1	1.5	
FT	427..461				1	1.5	
FT	/note= "Ser/Gly/Pro rich domain"				1	1.5	
FT	462..1147				1	1.5	
FT	/note= "C-terminal domain"				1	1.5	
PN	W09426922-A.				1	1.5	
PD	24-NOV-1994.				1	1.5	
PR	13-MAY-1994; UC5300.				1	1.5	
PR	01-MAY-1993; US-061697..				1	1.5	
PR	01-OCT-1993; US-131365.				1	1.5	
PA	(TEXA) UNIV TEXAS SYSTEM.				1	1.5	
PI	Briggs MR, Brown MS, Goldstein JL, Wang X;				1	1.5	
DR	WPI: 95-006813/01.				1	1.5	
DR	N-PSDB: 079037.				1	1.5	
PT	New sterol regulator element binding protein - used to develop				1	1.5	
PT	prods. and screening assays for agents for reducing plasma				1	1.5	
PT	cholesterol levels (Eng)				1	1.5	
PS	Claim 8; Page 185-98; 305pp; English.				1	1.5	
CC	The sequences given in R66390-91 represent the sterol regulatory element				1	1.5	
CC	binding proteins (SREBP), SREBP-1a and SREBP-2. SREBP's are regulatory				1	1.5	
CC	proteins which are involved in the regulation of genes involved in				1	1.5	

CC cholesterol metabolism that are under the control of an associated
 CC sterol regulatory element-1 (SRE-1) enhancer sequence. SREBP proteins
 CC fall into two families, SREBP-1 and SREBP-2. Both proteins are members
 CC of a family of basic-helix-loop-helix-leucine zipper (bHLH-Zip)
 CC transcription factors. Each have the ability to bind to SRE sequences
 CC and modulate SRE-mediated transcription. SRE-1 is a conditional enhancer
 CC found in the promoters for the low density lipoprotein (LDL) receptor
 CC and 3-hydroxy-3-methylglutaryl coenzyme A synthase genes. It increases
 CC transcription in the absence of sterols and is inactivated when sterols
 CC accumulate. Human SREBP-2 contains 1141 amino acids and has 47% identity
 CC with SREBP-1a. SREBP-1a was the first recognised member of this family
 CC and has 1147 amino acids. The resemblance between SREBP-1a and SREBP-2
 CC includes an acidic N-terminus, a highly conserved bHLH-Zip motif (71%
 CC identical), and an unusually long extension of 740 amino acids on the
 CC C-terminal side of the bHLH-Zip region. SREBP-2 possesses one feature
 CC lacking in SREBP-1a, a Glu rich region (27% Glu over 121 residues).
 CC SREBP promotes SRE-1-mediated gene transcription, eg. LDL receptor
 CC production in the presence of sterols. SREBP identified in screening
 CC assays, may be used to reduce plasma cholesterol levels and in
 CC controlling hypercholesterolaemia and its associated diseases.
 CC Sequence 1147 AA;

Query Match 18.6%; Score 1494; DB 1; Length 1147;

Best Local Similarity 33.6%; Pred. No. 4,82e-108; Mismatches 388; Conservative 227; Indels 55; Gaps 51;

Db 169 PGFSTGSPGNTQOPLPGLPLASPPGVPVSLHTQ-V-QSVVPQQLLTVAAPTAAPVT 226
 Oy 126 PTGLKAQPTATIHMDQRMPTNAVYPSLGSFVYQSMSPRTS-EVESANQNVNVM 184
 Db 227 TTVTSQIQOVPVLDPNHFKAQSLSLTAKTGATYKAGSLPYSVGTTCVQCPRLTLYS 286
 Oy 185 QPVAATPAPASAPRLPQOQSY-PQPFITYNSKA-GMTSDER-MYLLQPTVAPSPTPSPVAP 241
 Db 287 GGTITATVPLVYDAEKL-PI-NRLAAGKAPAS-AQSRG-E-KRTAHNAIEKRYSSIND 341
 Oy 242 PPTSTGSRASKVAVAPLAPSPAAEMVQGVNPNVOPKKEVKRSHNAIERRYRSTIND 301
 Db 342 KIIEKDLVVGTEAKLNKSAVLRAKADYIRFLOHNSNOKLQENSLT-RTAVHKSRS-LMD 399
 Oy 302 KINELKNLVYGQAKLNKSAVLKRSIDKTRDQROHDKAELOQRRLMARDSKYVD 361
 Db 400 LVS-ACGSGGNTDVLMEGKYT-EVEDTLTPPPSDAGSPQSSPFLSGSGSGSGSGSDS 457
 Oy 362 LLQGTTRPGRAKKRRESQGTFTTAGLTPRPSDESDPLS-PMH-SDLSLPPSPGSGT 419
 Db 458 EPDSPVFEESKAKPRDQRPFLSRGMLDRSLALCTVFLCTSCNPLASLGAAGLPSPD 517
 Oy 420 ASCSSGSSSSNEEPLVYPS-SMRGMAHTRSLGICMFMFAILAVNPFKTYL-QRGHYDSMD 477
 Db 518 TTSVYHSGRNVLTGTERDGPMAQWLLPVPVWLVNLGLVLSVLYFYGEPRVTRPHSG 577
 Oy 478 DLGDMSGQ-RRLTSTDYV-EGEGFAVWQSSWIMLNTLMGICVLYLYGGPQDQAQND 535
 Db 578 PAVYFWRRHKKQDLDLAKDFPQAQAQWLVA-LRALGRPLPTSHDLACSLL-WNLIN 634
 Oy 536 -A-YC-OHRORADFYSQGSQAAYAG-YLNCILHMGSLPLASRLE-CYLQTTQWQFLRF 589
 Db 635 LLQRLWVGWMLAGRAGGLQOQCALRYDASASARADALVYHKLHOLHTMKHNGHILTAN 694
 Oy 590 LFRHMLGVLVSRSGGLSNASARQALASARELLALNRLNQLQLTGSGRGDMNGIM 649
 Db 695 LALSALNTLAECGDAVSVATLAEIYVAALRYKTSLLPRALH-FLTFEFLSARQAC-LAQ 752
 Oy 650 MALFMSNAEVAHNLTPRETICIVMTALRMKRSAPKYLQDFPARYYSRAREGCGRR 709
 Db 733 SGSVAPAMQWLCHPYGHRFVDGDMVY-LSTP--WESLYSLAGNPYDPLAQTOLFRHL 809
 Oy 710 ATEQOELRMATFAYGYRCATHFVFLYDSDGEGDGFTRLRPNODPAAHYIKOYREHL 769
 Db 810 LERALNCTQPPSPGSGADGKEFSDALGYDLNLSGCSMAAAPAYSFSSSSM-ATTGG 868
 Oy 770 LFKSIQCLVAGAKHSGGLPTSSVSGEAEOLQOQOHSGITVSNVLYTKTSLKDTLWADDE 829

Db 869 VDPVAKMASITAVYIHWLRDEAEERLCLVEHLPRVIOE-SER-PLPRAALHS-FKA 925
 Oy 830 RDTNVVWMADEVLETAVHLLGEDTLAEOLYGRIKOMPTLOQCGENDILPKA-LHAVALRA 888
 Db 926 ARALL-GGAKA-ESGPALT-I-CEKAGYIQDSLATTPASSSID-KAV-QLFLCDLILY 979
 Oy 889 KMLIKNNMADLSLKLQVILDESSVELOECTVIRITDAGIKILFQLTCDWLE 948
 Db 960 VTSLSMROQPPAPAPAAQGSRRPQASALBGRFORDLSLRLAQSFPRAMRVEFLHE 1039
 Oy 949 TTTALM-ELEHMN-M-EDDGYQVP-GEVLE-K-FQTLNLSLRNVENIPNAQSRITYE 1002
 Db 1040 ATARLMAGASPTTHOLDRLRRRAGPGKG-GAVAELEPR--PTRREHAEALLIACY 1096
 Oy 1003 AVCRLMAGASCPMDQLDRSLRSRHAHSIFCGSKDRQCNFVGEGERERASAMYACKY 1062
 Db 1097 LPPGFLSAPGQRYGVLAAARTLEGLRRLIHLHQOQMLMGG 1141
 Oy 1063 LPPALLSSPGERAGMLAEAAKTLERKVKRLKCYOLMKSLGNG 1107

RESULT 2

ID R66391 standard; Protein; 1141 AA.

AC R66391.

DT 04-AUG-1995 (first entry)

DE Human SREBP-2.

KM Probe: sterol regulatory element binding protein; regulatory protein;
 KM cholesterol metabolism; sterol regulatory element-1; SREBP; SRE-1;
 KM SREBP-1; SREBP-2; basic-helix-loop-helix-leucine zipper; plasma;
 KM transcription factor; low density lipoprotein; LDL; receptor;
 KM 3-hydroxy-3-methylglutaryl coenzyme A synthase; cholesterol;
 KM hypercholesterolaemia; enhancer.

OS Homo sapiens.

FH Key

FT region

FT 1. 50

FT /note= "Acidic region"

FT domain

FT /note= "Ser/gly/Pro rich domain"

FT domain

FT /note= "Gln rich domain"

FT region

FT /label= bHLH-Zip

FT region

FT /note= "Basic region"

FT region

FT /note= "Helix 1"

FT region

FT /note= "Helix 2"

FT region

FT /note= "Putative leucine zipper"

FT domain

FT /note= "C-terminal domain"

PN W09426922-A.

PD 24-NOV-1994.

PF 13-MAY-1994; U05300.

PR 13-MAY-1993; US-061697.

PR 01-OCT-1993; US-131365.

PA (TEXA) UNIV TEXAS SYSTEM.

PI Briggs MR, Brown MS, Goldstein JL, Wang X;

DR WPI; 95-006813/01.

DR N-PSDB; Q79038.

PT New sterol regulator element binding protein - used to develop

PT prods. and screening assays for agents for reducing plasma

PT cholesterol levels (Eng)

PS Claim 10; Page 227-36; 305pp; English.

CC The sequences given in R66390-91 represent the sterol regulatory element

CC binding proteins (SREBP), SREBP-1a and SREBP-2. SREBP's are regulatory

CC proteins which are involved in the regulation of genes involved in

CC cholesterol metabolism that are under the control of an associated

CC sterol regulatory element-1 (SRE-1) enhancer sequence. SREBP proteins

CC fall into two families, SREBP-1 and SREBP-2. Both proteins are members

CC of a family of basic-helix-loop-helix-leucine zipper (bHLH-Zip)


```

Db      349  DNNLLERRRRRNNDRKEIGTLPKSSDDEMMKNGKITLKA5VDYIRKLQEQGRK- 407
Oy      286  SAINALERRRFRYSINDKINELKNLV--VGE-QAKLNSAVLRKSIDKIRDLQRONHDLKA 342
Db      408  DIESRQSL 416
Oy      343  ELQRLQREL 351

RESULT 4
ID      W52821 standard; Protein; 597 AA.
AC      W52821;
DE      08-SEP-1998 (first entry)
DE      Human PRCC-TFE3 construct protein from cell lines UOK120 and UOK146.
KW      PRCC; papillary renal cell carcinoma; TFE3; transcription factor;
KW      fusion protein; translocation; diagnosis; treatment.
OS      Homo sapiens.
OS      Synthetic.
FH      Key
FT      Protein
FT      Location/Qualifiers
FT      1..597
FT      /label= "PRCC-TFE3
FT      /note= "Fusion protein"
FT      Misc-difference 1..156
FT      /label= "PRCC
FT      /note= "papilloma renal cell virus partial sequence"
FT      Misc-difference 159..597
FT      /label= "TFE3
FT      /note= "transcription factor partial sequence"
PN      WO9806871-A1.
PD      19-FEB-1998.
PF      13-AUG-1997; G02209.
PI      13-AUG-1996; GB-016986.
PA      (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PI      Clark J, Cooper C, Shipley J;
DR      MPI; 98-159557/14.
DR      N-PSDB; V20957.
PT      Diagnosing papillary renal cell carcinoma by detecting gene
PT      trans-location - resulting in fusion of TFE3 gene with some other
PT      gene, also related vectors, transformed cells; specific binding
PT      reagents, peptide(s) encoded by fusions and therapeutic anti-sense
PT      sequences
PT      Claim 9; Fig 2B; 71pp; English.
CC      This sequence represents a novel fusion protein constructed from a
CC      papillary renal cell carcinoma (PRCC) associated protein and the
CC      transcription factor TFE3 which is found in cell lines UOK120 and UOK146
CC      and used in a method for the diagnosis, prophylactic and therapeutic
CC      treatment of papillary renal cell carcinoma. The translocation t(x;1)
CC      (p11.2;q21.2) found in PRCC results in a fusion of the TFE3 gene with a
CC      new chromosome 1 gene designated PRCC (at 1q21.2), resulting in
CC      expression of a fusion protein between the N-terminus of PRCC and almost
CC      the whole of the TFE3 gene. Normal TFE3 transcripts are no longer
CC      produced. Two other fusion partners for TFE3 have also been detected;
CC      Nono1 from a invx (p11.2; q13-24 or 12) translocation and the PSF splice
CC      factor gene, resulting in t(x;1) (p11.2;p34). These trans-locations
CC      define a subgroup of PRCC generally encountered in patients younger
CC      than 25.
SO      Sequence 597 AA:

Query Match 2.1%; Score 172; DB 1; Length 597;
Best Local Similarity 40.8%; Pred. No. 5,446-03;
Matches 29; Conservative 19; Mismatches 19; Indels 4; Gaps 3;

Db      326  KKNHNLERRRRFNNDRIKELGLTLPKSSDPEMRMKKGITLKA5VDYIRKLQEQGRS 385
Oy      284  KRAHAHALERRRFRYSINDKINELKNLV--VGE-QAKLNSAVLRKSIDKIRDLQRONHDL 340
Db      386  K-DIESRQSL 395
Oy      341  KALQRLQREL 351

```

ID	W52822: standard; Protein: 700 AA.
AC	M52822: (first entry)
DT	08-SEP-1998
DE	Human Nono/TFE3 fusion product from cell line UOK109.
DE	PRCC; papillary renal cell carcinoma; TFE3; transcription factor;
KW	fusion protein; translocation; diagnosis; treatment; Nono; p54-nrb.
OS	Homio sapiens.
OS	Synthetic.
FH	Key
FT	Protein
FT	Location/Qualifiers
FT	1..700
FT	/label= "Nono/TFE3"
FT	/note= "Fusion protein"
FT	1..377
FT	/label= "Nono"
FT	/note= "also known as p54-nrb"
FT	378..700
FT	/label= "TFE3"
FT	/note= "transcription factor sequence"
PN	WO9806871-A1.
PD	19-FEB-1998.
PF	13-AUG-1997; G02209.
PR	13-AUG-1996; GB-016986.
PA	(CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PI	Clark J, Cooper C, Shipley J;
PI	WPI: 98-15957/14.
DR	N-ESDB: V20956.
PT	Diagnosing papillary renal cell carcinoma by detecting gene
PT	trans-location - resulting in fusion of TFE3 gene with some other
PT	gene, also related vectors, transformed cells, specific binding
PT	reagents, peptide(s) encoded by fusions and therapeutic anti-sense
PT	sequences
PS	Claim 9; Fig 3A; 71pp; English.
CC	This sequence represents a novel fusion protein constructed from the Nono
CC	protein (also known as p54-nrb) and the transcription factor TFE3 which
CC	is found in cell line UOK109 and used in a method for the diagnosis,
CC	prophylactic and therapeutic treatment of papillary renal cell carcinoma.
CC	The translocation t(X;1)(p11.2;q21.2) found in papillary renal cell
CC	carcinoma (PRCC) associated protein (PRCC) results in a fusion of the
CC	TFE3 gene with a new chromosome 1 gene designated PRCC (at 1q21.2),
CC	resulting in expression of a fusion protein between the N-termini of
CC	PRCC and almost the whole of the TFE3 gene. Normal TFE3 transcripts are
CC	no longer produced. Two other fusion partners for TFE3 have also been
CC	detected; Nono, from a invx (p11.2;q13.24 or 12) translocation and the
CC	CC PSF splice factor gene, resulting in t(X;1)(p11.2;p34). These
CC	trans-locations define a subgroup of PRCC generally encountered in
CC	patients younger than 25.
SO	Sequence 700 AA:
Query Match	2.1%; Score 172; DB 1; Length 700;
Best Local Similarity	40.8%; Pred. No.5,44e-03;
Matches	29; Conservative 19; Mismatches 19; Indels 4; Gaps 3;
DB	429 KRONHLERRRRPNNIDIKELGTIPKSSDPEMMWNGKTIKASVDYIRKLOKEOORS 488
QY	284 KSAHAHAIRRRRTSINDKINELKNLV--VGE-QAKLNSAVLKSIDKIRLDIRONHDL 340
Db	489 K-DLESQRSL 498
QY	341 KAELORLQREL 351
RESULT	6
ID	W52820 standard; Protein: 834 AA.
AC	M52820:
DT	08-SEP-1998 (first entry)
DE	Human PRCC-TFE3 fusion protein from cell line UOK124.
KW	PRCC; papillary renal cell carcinoma; TFE3; transcription factor;
KW	fusion protein; translocation; diagnosis; treatment.
OS	Homio sapiens.
OS	Synthetic.
FH	Key
FT	Location/Qualifiers
FT	1..834
FT	/label= "PRCC-TFE3"

```
FT      MISC_difference 1..393
FT      /note= "fusion protein"
FT      /label= PRCc
FT      /note= "Papillary renal cell carcinoma partial
FT      sequence"
FT      MISC_difference 394..834
FT      /label= TFE3
FT      /note= "Transcription factor partial sequence"
PN      W09806871-A1.
PD      19-FEB-1998.
PR      13-AUG-1997: G02209.
PR      13-AUG-1996: GB-016985.
PA      (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PI      Clark J, Cooper C, Shipley J;
DR      WPJ; 98-159557/14.
DR      N-PDSB: V209556.
PT      Diagnosing papillary renal cell carcinoma by detecting gene
PT      translocation - resulting in fusion of TFEE3 gene with some other
PT      gene, also related vectors, transformed cells, specific binding
PT      reagents, peptides(s) encoded by fusions and therapeutic anti-sense
PT      sequences
PS      Claim 9: Fig 2A: 71pp: English.
CC      This sequence represents a novel fusion protein constructed from a
CC      papillary renal cell carcinoma (PRCc) associated protein and the
CC      transcription factor TFEE3 which is found in cell line UOK124 and used in
CC      a method for the diagnosis, prophylactic and therapeutic treatment of
CC      papillary renal cell carcinoma. The translocation t(X;1)(p11.2;q21.2)
CC      found in PRCC results in a fusion of the TFEE3 gene with a new chromosome
CC      1 gene designated PRCc (at i(21.2), resulting in expression of a fusion
CC      protein between the N-terminus of PRCc and almost the whole of the TFEE3
CC      gene. Normal TFEE3 transcripts are no longer produced. Two other fusion
CC      partners for TFEE3 have also been detected. NonO, from a invX(p11.2;
CC      q13-24 or 12) translocation and the PSF splice factor gene, resulting
CC      in t(X;1)(p11.2;p24). These translocations define a subgroup of PRCc
CC      generally encountered in patients younger than 25.
SQ      Sequence      834 AA:

Query Match          2.1%; Score 172; DB 1; Length 834;
Best Local Similarity 40.8%; Pred. No. 5.44e-03;
Matches    29; Conservative   19; Mismatches 19; Indels    4; Gaps    3;

Db       KKDHNHIERRRRINIDRIKELGTLPKSSDPEMNKGITIIKASVDYIRKILOEQR8 622
        | : |||||:::||::|||:|: : ::||::||::||::||::||::||::||::|:
Oy       284 KRSHNHLIERRYTRISINDKLNELNV--VGE-QAKLNKSAVLKSIDSIDRLDJRONHD 340
        | : |||::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy       341 KAELQRIQLREL 351

RESULT
ID      R62656 standard; Protein; 383 AA.
AC      R62656;
DT      08-JUN-1995 (first entry)
DE      Petunia ph6 gene product.
KW      vacuolar pH; pH gene; Petunia.
OS      Petunia hybrida strain Y26.
FH      Key location//Nullifiers
              192..198
FT      /label= Helix-loop-helix structural motif
FT      /note= "see also Z39..245"
```

[illegible]

CC The DNA, vectors and host cells of the invention are useful for the
 CC recombinant production of msin proteins useful in elucidation of Mad
 CC repressor functions.
 SO Sequence 160 AA:

Query Match 1.9%; Score 149; DB 1; Length 160;
 Best Local Similarity 29.9%; Pred. No. 1.83e-01;
 Matches 26; Conservative 28; Mismatches 30; Indels 3; Gaps 3;

DB 7 IEVESDEDEGRPOSAA-D-KRAHNALERRRDKHDKSFHSIDSVSLQGEKASRAQIL 64
 QY 265 MEYQGVKVINRVOQPKYKVEVRSANALERRRYRISINDKINELKNLVGEQA-KLNKSAVL 323
 DB 65 DKATEYIOYMRKRNHTHQDIDDLKRO 91
 QY 324 RKSIDKIRDLQRONHDKAKELQRIQRE 350

RESULT 9
 ID R84882 standard; Protein: 2414 AA.
 AC R84882;
 DT 01-FEB-1996 (first entry)
 DE Transcription factor p300.
 KW Transcription factor; p300; adenovirus; early region 1A; E1A;
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT peptide 11..17
 FT region /label= Nuclear_location_signal
 FT 342..421
 FT /label= C/H-rich_region_1
 FT /note= "cysteine/histidine-rich region containing
 FT 2 putative zinc finger motifs"
 FT domain 1070..1134
 FT /label= Bromodomain
 FT region 1162..1461
 FT /label= C/H-rich_region_2
 FT 1622..1821
 FT /label= C/H-rich_region_3
 PN WO9528499-A1.
 PD 26-OCT-1995.
 PF 13-APR-1995; U04682.
 PR 14-APR-1994; US-227336.
 PA (DAND) DANA FARBER CANCER INST INC.
 PI Eckner R, Ewen M, Livingston D;
 DR WPI: 95-373813/48.
 DR N-PSDB: T02792.
 PT Nucleic acid encoding human p300 that associates with adenovirus E1A
 PT - and related vectors, host cells and screening assays, also
 PT diagnosis of cancerous and pre-cancerous tissue by detection of
 PT mutant p300
 PS disclosure; Page 61-78; 126pp; English.
 CC Transcription factor p300 was isolated from human 293 cells and used
 CC to generate polyclonal antisera in mice. These were used to screen
 CC 293 CDNA libraries to isolate clones contg. overlapping inserts,
 CC which were assembled to obtain a full-length CDNA sequence (T02792)
 CC encoding a protein (R84882) of predicted mol. wt. 264.236 kDa.
 CC p300 may be produced in host cells (pref. mammalian) and used to
 CC raise monoclonal antibodies or to screen cpds. for the ability to
 CC modulate p300-dependent transcription.
 SO Sequence 2414 AA:

Query Match 1.9%; Score 149; DB 1; Length 2414;
 Best Local Similarity 24.9%; Pred. No. 1.83e-01;
 Matches 48; Conservative 50; Mismatches 81; Indels 14; Gaps 11;

DB 818 HCFQLPQALHONSPPVSRPTPTPHHTPPSIGAQQP-ATTIPAVPTPPAMPQPSQ 876
 QY 120 NCPQOQPTGLKKAQPTATIHMDAQRMPNTAVYPPSLGSSFYQSMSPPTSPYESAQ 179
 DB 877 ALHP-PP-ROTPPTPTQLQOV--QPSLPAAPSADQPOQPRSQ--QSTAAVSPTPNAP 930
 QY 180 NVNVMQPVAAATPAASAPLPQOQSYPOPFITYNSKAGMTSDAMTLLQPTVAS-PPPSPP 238

DB 931 ILPPQATPLSPQAVSIEGVSNPSTSTEVNSQ-AIAEKOPS-OEVKMEAKMEYDQE 988
 QY 239 VAPPTSTGSRASKVAV-APLA--PSPAMEVGKVPINRVOQPKYKVEVRSANALERRY 295

DB 989 PADTQPEDISESK 1001
 QY 296 RTSIN-DKINELK 307

RESULT 10
 ID W40057 standard; Protein: 2414 AA.
 AC W40057;
 DT 20-JUL-1998 (first entry)
 DE Cellular transcriptional factor p300.
 KW Cellular transcriptional factor; p300; human; p300; P/CAF;
 KW transcription; histone acetyltransferase; HIV; infection; cancer;
 KW therapy; muscle differentiation.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Region 1763..1966
 FT /note= "P/CAF binding region"
 PN WO9803652-A2.
 PD 29-JAN-1998.
 PF 23-JUL-1997; U12877.
 PR 23-JUL-1996; US-022273.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Howard BH, Nakatani Y;
 DR WPI: 98-120777/11.
 DR N-PSDB: V10092.
 PT New isolated p300/CBP-associated factor, P/CAF - used to develop
 PT products for modulating transcription, e.g. for treating HIV
 PT infection or cancers or for promoting muscle differentiation
 PS disclosure; Page 76-81; 107pp; English.
 CC This polypeptide sequence comprises p300, a global transcriptional
 CC coactivator that is involved in the regulation of various
 CC DNA-binding transcriptional factors. The invention relates to a
 CC novel human p300/CBP associated cofactor, P/CAF (see W40052), that
 CC modulates transcription through binding to p300 and CBP (see
 CC W40058). The region (see W40055) of p300 that binds to P/CAF
 CC is claimed. The invention provides methods of screening for
 CC compounds that inhibit or stimulate the transcription modulating
 CC and histone acetyltransferase activity of P/CAF and p300/CBP.
 CC Inhibitors can be used e.g. to inhibit HIV TAR-mediated
 CC transcription in the treatment of HIV infection. Stimulators can
 CC be used e.g. to activate tumour suppressor p53 in the treatment of
 CC cancer or to activate the muscle differentiation factor MyoD to
 CC promote muscle differentiation. The products can also be used to
 CC inhibit the cell cycle progression inducing effect of an
 CC oncoprotein which binds p300/CBP in a subject. Also provided is
 CC a method for determining the amount of P/CAF in a sample by
 CC contacting the sample with the P/CAF binding region of p300 and
 CC determining the amount of P/CAF/p300 complex formed.
 SO Sequence 2414 AA:

Query Match 1.9%; Score 149; DB 1; Length 2414;
 Best Local Similarity 24.9%; Pred. No. 1.83e-01;
 Matches 48; Conservative 50; Mismatches 81; Indels 14; Gaps 11;

DB 818 HCFQLPQALHONSPPVSRPTPTPHHTPPSIGAQQP-ATTIPAVPTPPAMPQPSQ 876
 QY 120 NCPQOQPTGLKKAQPTATIHMDAQRMPNTAVYPPSLGSSFYQSMSPPTSPYESAQ 179
 DB 877 ALHP-PP-ROTPPTPTQLQOV--QPSLPAAPSADQPOQPRSQ--QSTAAVSPTPNAP 930
 QY 180 NVNVMQPVAAATPAASAPLPQOQSYPOPFITYNSKAGMTSDAMTLLQPTVAS-PPPSPP 238
 DB 931 ILPPQATPLSPQAVSIEGVSNPSTSTEVNSQ-AIAEKOPS-OEVKMEAKMEYDQE 988
 QY 239 VAPPTSTGSRASKVAV-APLA--PSPAMEVGKVPINRVOQPKYKVEVRSANALERRY 295
 DB 989 PADTQPEDISESK 1001

OY 296 RTSIN-DKINELK 307

RESULT 11
ID M10042 standard: Protein: 151 AA.
AC M10042:
DT 07-FEB-1998 (first entry)
DE Max 14.
KW murine; mSinA: mammalian homologue; Saccharomyces cerevisiae: repressor;
KW Sin3; Mad; Max: mSin:Mad complex; mSin:Mad complex; Myc: promoter;
KW basic helix-loop-helix zipper protein; compete: DNA-binding;
KW Myc:Max complex; activate: transcription; gene regulation.
OS Homo sapiens.
PN US624818-A.
PD 29-APR-1997.
PF 01-JUN-1994: 252966.
PR 01-JUN-1994: US-252966.
PR 19-SEP-1991: US-756195.
PR 23-JUN-1992: US-903710.
PR 01-APR-1994: US-222638.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI Ayer DE, Eisenman RN;
DR WPI: 97-258216/23.
DR N-PSDB: T70134.
PT mSin nucleic acids encoding recombinant polypeptide(s) that
PT associate with Mad polypeptide - are possible homologues of S.
PT cerevisiae general repressor protein
PS Example 2: Fig 2D; 11pp; English.
CC This sequence represents Max 14, a basic helix-loop-helix zipper
CC (bHLHZip) protein. Max 11 and Max 14 cDNAs appear to be partial,
CC overlapping cDNAs. Subsequent isolation of several overlapping cDNAs
CC from a Manca (human Burkitt's lymphoma cell line) lambda gt10 library
CC permitted deduction of an apparently complete open reading frame for
CC Max that encodes 151 residues. The 9-amino acid insertion found in
CC several PCR clones is not indicated in the specification. Max is an
CC obligate partner for the DNA binding and transcriptional functions of
CC Myc family proteins as well as for the Mad protein. Max is a stable,
CC ubiquitously expressed protein which in general does not appear to be
CC regulated during mitogenesis, the cell cycle, or differentiation.
CC Expression of Mad is closely linked to differentiation in at least two
CC distinct cell lineages. The switch from Myc:Max to Mad:Max complexes may
CC reflect the repression of transcription of Myc regulated genes by Mad.
CC The DNA, vectors and host cells of the invention are useful for the
CC recombinant production of mSin proteins useful in elucidation of Mad
CC repressor functions.
SQ Sequence 151 AA:

Query Match 1.8%; Score 148; DB 1; Length 151;
Best Local Similarity 32.4%; Pred. No. 2.13e-01;
Matches 22; Conservative 23; Mismatches 22; Indels 1; Gaps 1;

Db 15 KRAHNALERRRDIKSFSLRDSVPSLOGEKASRAQIIDKATEYIYRKRNHTHQO 74
OY 284 KRSAHNALEERRYRTSINKINELKNLVGEQA-KLNKSAVLKRSIDKTRDLOPNHDLKA 342
Db 75 DIDLKRRQ 82
OY 343 ELQRLQRE 350

RESULT 12
ID R33386 standard: Protein: 160 AA.
AC R33386:
DT 15-JUL-1993 (first entry)
DE Max protein.
KW Mad; max; myc; C-myc; helix-loop-helix zipper; leucine zipper;
KW helix-turn-helix; diagnosis; prognosis; cancer; malignancy;
KW neoplasm; tumour; studying embryogenesis; study gene regulation.
FH Key Location/Qualifiers
FT region 9..17
FT /note= "9 amino acid insertion found in several
FT PCR clones."
FT domain 39..50

FT 1 /label= Helix I
FT /note= "Helix I of b-HLH homology region."
FT 57..75
FT domain /label= Helix II
FT /note= "Helix II of b-HLH homology region"

PN W09305056-A.
PD 18-MAR-1993.
PF 09-SEP-1992: U07629.
PR 09-SEP-1991: US-756195.
PR 23-JUN-1992: US-903710.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI Ayer DE, Blackwood EM, Eisenman R;
DR WPI: 93-100913/12.
DR N-PSDB: Q38720.
PT Helix-loop-helix zipper protein named Max - associates with Myc
PT or Mad polypeptide(s), useful as diagnostic or prognostic tools
PT for diverse types of cancer
PS Claim 3; Fig 2; 93pp; English.
CC This sequence represents the Max protein, and was decoded from the
CC cDNA isolated as in Q38720. A computer search on a protein database
CC revealed sequence similarity between a segment of the Max ORF and
CC the b-HLH proteins, including members of the Myc family. The Max
CC sequence in this region represents a nearly exact match with the
CC HLH consensus. This similarity also extends in the amino-terminal
CC direction into a basic region of Max. The Max sequence just carboxyl
CC terminal to helix II contains a series of hydrophobic residues,
CC three of which are leucines, spaced seven residues apart. Helical wheel
CC analysis of this region suggests that the amphipathic helix II may
CC extend into and beyond the three leucines. These leucines and the
CC other non-polar residues might form a hydrophobic face similar to
CC that in the leucine zipper proteins. It was also shown that a 15
CC amino acid deletion of the basic region abolishes the capacity of
CC c-myc to co-transform Rat-1 cells in collaboration with the bcr-abl
CC oncogene.
SQ Sequence 160 AA:

Query Match 1.7%; Score 140; DB 1; Length 160;
Best Local Similarity 32.4%; Pred. No. 5.99e-01;
Matches 22; Conservative 22; Mismatches 23; Indels 1; Gaps 1;

Db 24 KRAHNALERRRDIKSFSLRDSVPSLOGEKASRAQIIDKATEYIYRKRNHTHQO 83
OY 284 KRSAHNALEERRYRTSINKINELKNLVGEQA-KLNKSAVLKRSIDKTRDLOPNHDLKA 342
Db 84 DIDLKRRQ 91
OY 343 ELQRLQRE 350

RESULT 13
ID Y07044 standard: Protein: 258 AA.
AC Y07044:
DT 02-JUL-1999 (first entry)
DE Breast cancer associated antigen precursor sequence.
KW Breast cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
OS Homo sapiens.
PN W09904265-A2.
PD 28-JAN-1999.
PF 15-JUL-1998: U14679.
PR 22-JUN-1998: US-102322.
PR 17-JUL-1997: US-896164.
PR 10-OCT-1997: US-061599.
PR 10-OCT-1997: US-061765.
PR 10-OCT-1997: US-948705.
PR 11-OCT-1997: GH-021697.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ,
PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E,
PI Tureci O;
DR WPI: 99-132448/11.
PT New isolated cancer associated nucleic acids and polypeptides -

PT Isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
PS Disclosure: Page 428-429; 787pp: English.
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
SQ Sequence 258 AA;

Query Match 1.7%; Score 135; DB 1; Length 258;
Best Local Similarity 31.9%; Pred. No. 1.45e+00;
Matches 23; Conservative 22; Mismatches 23; Indels 4; Gaps 4;

Db 13 DPHGLTEETHOGRKINA-REASQTEKRRRDKNSFIDELASLYPTCNAMSRKIDKLTIV 71
OY 266 EVOGKVPINRQPKYKEVRSANNAIERRYRTSINDKINELKNLV-V-GEQA-KLNKSAV 322
Db 72 LRMAVOHMKTLR 83
OY 323 LRKSIDKINDLO 334

RESULT 14
ID R38751 standard; Protein; 644 AA.
AC R38751:
DE 23-DEC-1993 (first entry)
DE DEL protein.
KW DEL; snapdragon A. majus; regulation; pigmentation; pattern; ivory;
KW anthocyanin; del mutation; phenotype; corolla; lobe; transformation;
KW LC: R-S: R gene family; maize; expression vector; marker.
OS Antirrhinum majus.
PN WO9314211-A.
PD 22-JUL-1993.
PF 08-JAN-1993; G000019.
PR 09-JAN-1992; US-818570.
PA (CARP/) CARPENTER R.
PA (COEN/) COEN E.
PA (FREE/) FREEDMAN R.
PA (GOOD/) GOODRICH J.
PI Carpenter R, Coen E, Freedman R, Goodrich J;
DR WPI: 93-243226/30.
DR N-PSDB: 046554.
PT DNA mol. encoding DEL protein or homologue - capable of
PT regulating expression of plant genes involved in pigmentation
PT biosynthesis for regulating anthocyanin genes and plant
PT pigmentation
PS Claim 1: Page 26-29; 61pp: English.
CC This sequence represents the DEL protein derived from the snapdragon,
CC Antirrhinum majus, which is involved in pigmentation pattern
CC genes. The recessive del mutation causes a marked difference in
CC phenotype, causing corolla tubes to be ivory and the lobes to be
CC fully pigmented. The DEL protein has a strong homology to the
CC products of lc and R-S, two members of the R gene family which
CC controls pigmentation in maize. The DNA sequence encoding this
CC protein may be used, in an expression vector, to control pigmentation
CC in a range of plants. The DEL coding region may also be used as a
CC visible marker for gene expression for the detection of transformed
CC cells.
SQ Sequence 644 AA;

Query Match 1.7%; Score 134; DB 1; Length 644;
Best Local Similarity 31.8%; Pred. No. 1.68e+00;

Matches 21; Conservative 19; Mismatches 24; Indels 2; Gaps 2;
Db 435 PTAAEDIRN-VLSEKREKINERFMILASLVPSG-GKVQVSIIDHTIDYLGLEKRV 492
OY 278 PKYKEVRSANNAIERRYRTSINDKINELKNLVYGEQAKLNKSAVLRKSIDKINDLOKRN 337
Db 493 DELESN 498
OY 338 HDLKA 343

RESULT 15
ID W52830 standard; Protein; 1004 AA.
AC W52830:
DE 09-JUL-1998 (first entry)
DE A tumour suppressor gene called polyhomeotic 1 (hpl1).
KW Tumour suppressor gene; polyhomeotic 1; hpl1; human chromosome 12p13;
KW identification; neoplastic tissue; cellular differentiation; diagnosis;
KW neoplasia; ss.
OS Homo sapiens.
PN WO9807858-A1.
PD 26-FEB-1998.
PF 22-AUG-1997; U14866.
PR 06-FEB-1997; US-036939.
PR 23-AUG-1996; US-024349.
PR 04-DEC-1996; US-031569.
PA (CHIR) CHIRON CORP.
PI Randazzo F;
DR WPI: 98-169162/15.
DR N-PSDB: V21060.
PT Isolated human polyhomeotic 1 oncogene - used to develop products
PT for diagnosis and therapy of proliferative and developmental
PT disorders, e.g. neoplasia, dysplasia or hyperplasia
PS Claim 1: Pages 28-30; 40pp: English.
CC The present sequence encodes a novel human tumour suppressor gene
CC termed polyhomeotic 1 (hpl1). The hpl1 gene maps to human chromosome
CC 12p13, a region which is frequently lost in non-small cell lung cancer
CC and breast cancer. A method of identifying neoplastic tissue of a
CC human comprises comparing the expression of a hpl1 gene in a tissue of
CC a human suspected of being neoplastic with the expression of a hpl1 gene
CC in a tissue of the human which is normal. Under-expression of the hpl1
CC gene identifies the subject as having neoplastic tissue. The hpl1
CC oncogene functions to suppress neoplasia and dysplastic or hyperplastic
CC cell growth as well as to induce cellular differentiation. The cDNA,
CC protein and vectors can be used as diagnostic and therapeutic tools for
CC proliferative and developmental disorders and to identify a p13 region
CC of a human chromosome 12. They can be used for the detection, diagnosis
CC or prognosis of neoplasia or for detecting a genetic predisposition to
CC neoplasia. They can also be used to treat tumours.
SQ Sequence 1004 AA;

Query Match 1.7%; Score 133; DB 1; Length 1004;
Best Local Similarity 20.2%; Pred. No. 1.94e+00;
Matches 49; Conservative 56; Mismatches 123; Indels 12; Gaps 12;

Db 372 TOIOPHSLLIOOQOQIHLQOKOVYIOQOIAIHQOQFQHQSOQLTATHTALQAOQOQOQ 431
OY 48 APTQWYNMLDEPRHTHTQOQSVDDQPSVEQPHVKSEHSVHI-KF-ELHQOQOQSP 105
Db 432 QOQOQOQOQOATTLTAPPOVPYPPQOVPSSQOQOATLVVQPMQSSPLSPD-A-A 489
OY 106 LLVYRPDLIAISYNCPOQOQPTGLKAAQPTATIHMDAORPMTAVYPPSLGSSFYVQ 165
Db 490 PK-PPI-PIOSKPPVAPLIPKPOLGAKKMSAQQRPPIHVPQVYG-TROPGTAAQALGLA 546
OY 166 SMSPTSPVESNONVWVNOVPAAPAPASAPLPQOQTPPEFTITYSKAGMTSDEAMTLL 225
Db 547 QUAAVPTSRKPGTGVOSGAHILASSPSSQAPGALQECPTLA-PG-MTIAVYOGTAHV 604
OY 226 -LQPTVASPTSP-PVAPPTSTGSRASKVRVA-PLASPAMEVQKVPINRQPKYKE 282
Db 605 VK 606
||

Tue Sep 5 10:30:02 2000

US-09-332-522B-8.rag

Page 9

OY 283 WK 284

Search completed: Sun Sep 3 12:54:06 2000
Job time : 54 secs.

THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

W E S T
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSEARCH protein - protein database search, using Smith-Waterman algorithm
Run on: Sun Sep 3 12:51:03 2000; Maspar time 53.56 Seconds
Tabular output not generated. 980.293 Million cell updates/sec

Title: >US-09-332-522B-8
Description: (1-1113) from US09332522B.pep
Perfect Score: 8012
Sequence: 1 MDITLMLNIDAPLDESMDLF.....LKECYQLMKSLGNGISYKA 1113

Scoring table:
PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p164
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 56.181; Variance 139.522; scale 0.403

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	8012	100.0	1113	2 T13348	transcription factor	0.00e+00
2	1418	17.7	1133	2 A54164	steroid regulatory ele	1.81e-204
3	1328	16.6	1141	2 A54962	steroid regulatory ele	2.41e-189
4	1282	16.0	1139	2 B54962	steroid regulatory ele	1.25e-181
5	1001	12.5	927	2 A48085	transcription factor	8.29e-135
6	336	4.2	391	2 PD0035	steroid regulatory ele	1.03e-28
7	192	2.4	446	2 A42029	transcription factor	1.01e-08
8	182	2.3	419	2 T38024	MTF protein - human	1.81e-07
9	181	2.3	419	2 A40728	microphthalmia-associ	2.41e-07
10	182	2.3	520	2 T14752	microphthalmia-associ	1.81e-07
11	172	2.1	536	2 A34596	transcription factor	3.06e-06
12	162	2.0	514	2 A35658	transcription factor	4.83e-05
13	159	2.0	685	2 T04073	intensifier 1 protein	1.09e-04
14	149	1.9	103	2 A42611	basic-helix-loop-heli	1.55e-03
15	149	1.9	134	2 S33120	max protein 3 - human	1.55e-03
16	149	1.9	150	2 S33118	max protein 1 - human	1.55e-03
17	149	1.9	160	2 A38488	Myb protein - mouse	1.55e-03
18	149	1.9	160	2 T50379	gene max protein - ch	1.55e-03
19	150	1.9	317	2 A48080	basic helix-loop-heli	1.19e-03
20	151	1.9	474	2 S15921	protein TPX-VT3 - The	9.17e-04
21	150	1.9	817	2 S51342	veripolin - yeast (Sa	1.19e-03
22	149	1.9	2414	2 A54277	transcription adaptor	1.55e-03
23	148	1.8	136	2 T51585	Myc binding protein -	2.01e-03

Query Match	Best Local Similarity	100.0%	Score 8012	DB 2	Length 1113	Pred. No. 0.00e+00	Mismatches 0	Indels 0	Gaps 0
Matches 1113	Conservative	0	Mismatches	0	Indels	0	Gaps	0	

ALIGNMENTS

RESULT 1

ENTRY T13348 #type complete
TITLE Transcription factor HbH106 - fruit fly (Drosophila melanogaster)ORGANISM #formal_name Drosophila melanogaster
DATE 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 20-Sep-1999ACCESSIONS T13348
REFERENCE Z17656
#authors Theopold, U.; Ekengren, S.; Hultmark, D.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:1195-1199
#title HbH106, a Drosophila transcription factor with similarity to the vertebrate steroid responsive element binding protein.
#cross-references: MIM:9616543#accession T13348
#status preliminary: translated from GB/EMBL/DBJ
#molecule_type mRNA
##residues 1-1113 #label THE
##cross-references EMBL:U38238; NID:g1079655; PID:g1079656;CLASSIFICATION #superfamily steroid regulatory element binding protein
SUMMARY #length 1113 #molecular-weight 124513 #checksum 539

Query Match	Best Local Similarity	100.0%	Score 8012	DB 2	Length 1113	Pred. No. 0.00e+00	Mismatches 0	Indels 0	Gaps 0
Matches 1113	Conservative	0	Mismatches	0	Indels	0	Gaps	0	

D	b		301	DKINELKLVGEBOAKLNKSAVLRKSIDTKIRDLQRONHDLKAFLQRLQRELMARODSKVK	360
Q	y		301	DKINELKNLVYEBQAKLNKSAVLRKSIDKIRDLQRONHDLKAELQRLQRELMARADGSKVK	360
D	b		361	DLLOLGTPRGRASKKRRESSQTFTTAGLTTPPSDDSDSPSLSMHSDISLPSPYGSTA	420
O	y		361	DLLOLGTFRGRASKKRRESSQTFTTAGLTTPPSDESDPSLSMHSDISLPSPYGGSTA	420
D	b		421	SCSSGSSSSNEEPLVYPSSMRGMATHSRGLCMFMFAILAVNPFKTFLRGHYDSNDLG	480
O	y		421	SCSSGSSSSNEEPLVYPSSMRGMATHSRLCLCMFMFAILLAVNPFKTFLRGHYDSNDLG	480
D	b		481	DMSGORLLISTYVEEGGFVAWQOSSIMTLNFLMLGCLYKLLYVGDPDQDAOTDACQH	540
O	y		481	DMSGORRLISTYVEEGGFVAWQOSSIMWLNLFLMLGCLYKLLYVGDPDQDAOTDACQH	540
D	b		541	RORDPYFSOGOSQAAYAGLNCILHNEGSLPASPRLCYQTWTWOFLRFPHRLMIGRYLV	600
O	y		541	RORDPYFSOGOSQAAYAGLNCILHNFGSLPASPRLCYQTWTWOFLRFPHRLMIGRYLV	600
D	b		601	SRRSGGLFESMAASRKQALASARELALLFNRLNLOJLTGNGSRGDMMGIMMALPASNAAEV	660
O	y		601	SRRSGGLFESMAASRKQALASARELALLFNRLNLOJLTGNGSRGDMMGIMMALPASNAAEV	660
D	b		661	AANLLTPRETICIHVTALTALRMKRSARKYLQOFPARYMTSAROBCEGTATTEQTQELRWA	720
O	y		661	AANLLTPRETICIHVTALTALRMKRSARKYLQOFPARYMSAROBCEGTATEQTQELRWA	720
D	b		721	FPAAYGRYCATHFTYDLDSPGEDGFETRLRPDCPAAHVIKOYREHLFFKSIQCLVGA	780
O	y		721	FPAAYGRYCATHFTYDLDSDSGEDGFETRLRPDCPAAHVIKOYRHLLFFKSIQCLVGA	780
D	b		781	GHRSGGLPTSSVSGBEAEOLOQQOHSGITVSNVLKYTSLLKDITMADEDERTNVVMADV	840
O	y		781	GHRSGGLPTSSVSGBEAEOLOQQOHSGITVSNVLKYTSLLKDITMADEDERTNVVMADV	840
D	b		841	LETAHWMLIGEDTLAOLQYGRITKOMPTOLOQCGENHLPALAHVYLRAXMILLKNNGNAL	900
O	y		841	LETAHWMLIGEDTLAOLQYGRITKOMPTOLOQCGENHLPALAHVYLRAXMILLKNNGNAL	900
D	b		901	DKSLKQLVNILOESSVELOECITVNRIIPAKGIKILFOLLTCDMLETFTALMEIHMN	960
O	y		901	DKSLKQLVNILOESSVELOECITVNIITDAKGILKLFOLLTCDMLETFTALMEIHMN	960
D	b		961	MEDDGFYQVPAGEYLEFEQTDILNSLRNIVENIPAASRIYLEYEAVCRIMAGASPCPTOQL	1020
O	y		961	MEDDGFYQVPAGEYLEFEQTDILNSLRNIVENIPAASRIYLEYEAVCRIMAGASPCPTOQL	1020
D	b		1021	DPSLSRNHAHSSIFCGSKDRROONFVGGEEPERASAMTVACKYLPALLSSGPFGAGLAIE	1080
O	y		1021	DPSLSRNHAHSSIFCGSKDRROONFVGGEEPERASAMTVACKYLPALLSSGPFGAGLAIE	1080
D	b		1081	AAKTELEVGDKRKLKECYOLMKSLGNGISYKA	1113
O	y		1081	AAKTELEVGDKRKLKECYOLMKSLGNGISYKA	1113

RESULT 2

ENTRY A54164 #type complete

TITLE Sterol regulatory element-binding protein 1 - Chinese hamster

ORGANISM #Normal.name Citricellus griseus #common.name Chinese hamster

DATE 07-Jul-1995 #sequence.revision 07-Jul-1995 #text.change 24-Sep-1999

ACCESSIONS A54164

REFERENCE A54164

#authors Scto, R.; Yang, J.; Wang, X.; Evans, M.J.; Ho, Y.K.; Goldstein, J.L.; Brown, M.S.

#journal J. Biol. Chem. (1994) 269:17267-17273

#title Assignment of the membrane attachment, DNA binding, and transcriptional activation domains of sterol regulatory element-binding protein-1 (SREBP-1).

#cross-references M0ID:94274723

#accession A54164

	##status	preliminary
	##molecule-type	mRNA
	##residues	1-1133 ##label SAT
	##cross-references	GB:009103; NID:9516002; PIDN:AAA20085.1; PID:9516003
	CLASSIFICATION	#superfamily 103; #regulatory element binding protein
	SUMMARY	#length 1133 #molecular-weight 120464 #checksum 4586
Query Match	17.7%;	Score 1418; DB 2; Length 1133;
Best Local Similarity	32.8%;	Pred. No. 1,81e-204;
Matches 324;	Conservative 240;	Mismatches 364; Indels 59; Gaps 52.
Db	174	LPGNQGPSPSLSASA-PGVSP-IS-LHPOVSSASQDPLPASTAPRTTTSQIOIRVP 230
QY	147	MEPNTAVTPSLGSSFTYQSMSPPTSPYESNQNQNVNQPAATRAPASAPL-PQ-QSYR 204
Db	231	VLPQHFIFKADSLLTYYTKTDTGATMTKAGISTLAPGTAVOAGPLQTLVSGGTLATVPL 290
QY	205	--QF-FITYSKSGMT-SDEAMLLLOPTVASPTSPSPVAPPTSTSGSRASKRYVA-PL 258
Db	291	VVDTDKLPRIHLAAGSKALGSAOSRGEKRTAHNAIEKRYNSINDKLYELKDLVVGTEAK 350
QY	259	APSPAMAEVQGVKPINR-V-OPKKEVRSAHNAIEREYRISINDKINELNLVVGQAK 316
Db	351	LKNSAVLEKRAIDYRFLQHSNOKLKOEMAL-RN-AAHKSRLDVLVYASGASG-GRDVA 407
QY	317	LKNSAVLEKRSIDKTRDLORONHDKAEQRLQRELEMAFDGSKYVDLLOLGRPRASKKR 376
Db	408	MEGVPEVVDTR-LPPPSDAGSPQSSPPLS-LGSRGSSSGGSDSE-PDSPYEDSOYKAQ 464
QY	377	RSSQTFPTTDAGLPFRPSDESDPLSRPHSDISLPRPYGGSTASCSSGSSSNEEPLVY 436
Db	465	RLHSGMLDRSRALCALVFLCTICNPRLPSFGMGIEPSSASAHHSQSMLEAESRD 524
QY	437	PSSMGMAITHSLRCMFMFAILVNPETFLQGHVSDNDLDM-SGQRRIISTUYV-E 494
Db	525	GSNNQWLLPRLVWLANLVLVACIALFLFYGEVTPRHTSPRAVHNRKQADLDLARG 584
QY	495	GGGFVWVOOSKIMLNTLTMIGCIVKLYLVGDQDLADQD-A--YCONHRADPFYTSQG 551
Db	585	DEPAQAQDLMTLA-LQALGRPLPTSLNDLACSL--WNILRIHLQRLVWGMVLAGRGLR 641
QY	552	QSSQAVAG-YLNCMLHPELSIPASLE--CYLQTTWQRLRLFRMLWGRVLSRSGGLE 608
Db	642	RDCGLRMDARASARDALVYHKLHOLHMGKYYTGGHILASVLAISALNLAECAGDAVS-M 700
QY	609	SNAAROKLISARELALLENRLNLOLDTGSGSDMGIMMALEFASMAEVAHANLTPR 668
Db	701	ALALATYYAALIRVTSILPRALH-ELTFEFLISARQC-LAOSGSVPLAMOWLCHPRGHR 758
QY	669	ETTC-IHWMTALRMKRSAPKWLQGFARYYSRAQOBGRTIRAEQTOELMWAFTATGYR 727
Db	759	F-FVVG-DMAV-HGARQESLISVAGNPVDPPLAQTRLFCEHILERRALNCIAPSPGTAD- 814
QY	728	YCAHVFYTYDLSDESGEGFFRLRNPQDPAHAYIKQYREHLLKRSICDLVGAHKKSGGL 787
Db	815	GDREFSDALGYQLINRCSDAVGTPA-CSFSVSSM-ASTTGDPVAKWASLTVAVVIHW 872
QY	788	PTSSVSGAEQLOOQHSGTIVSNLYKTSILKDLTIMADEDEBDPTVWMAVDLETAIVHW 847
Db	873	LARDEAARLPLVERRPHVLOE-TER-PLPKALYS-FKAARALLDHR-K-VESGPAS 927
QY	848	LIGEPTLEAQLGRKOMPLOLOOGENDHLPKA-IHAVLRAKMILKKNNSNALDKSLQK 906
Db	928	LA-I-CENASGAGRLSRLAAPTGGSID-KAM-OLLCOLLVARTSMQROSPASAOVA 983
QY	907	LNNITCDESSVELQCLTVNRTTDAKGIKLLFOLLTCMWLETRIALWELEHMMWEDGF 966
Db	984	HSASNGSQAFALELGFQODLSLRLKQNERPPAMRVFLTHEATARLWAGASPARTHOLL 1043
QY	967	YQV-P-G-EV--LE-K-FQTDLSLNAYENIPNQSRIITLEVACRLMAGASPCPTQQL 1020
Db	1044	DRSLRRRAGSSGKG-GTVAELPR-PTWREHTEALLASCLDPPLATFADGQOMSLAE 1100

QY 1021 DRLSRNAHSSIFCGSDRQONFVGGERASAMVACKYLPALLSPGERAGMLAE 1080

Db 1101 AARTVEKLGDRHLLDCCOMLLRLGG 1127

QY 1081 AAKLEKVGDRKRLKECYQLKMSLNG 1107

RESULT 3

ENTRY A54962 #type complete

TITLE sterol regulatory element binding protein 2 precursor - human

ALTRNAME_NAMES basic-helix-loop-helix-leucine zipper transcription factor

ORGANISM #formal_name Homo sapiens #common_name man

DATE 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 24-Sep-1999

ACCESSIONS A49397 A54962

REFERENCE #authors Hua, X.; Yokoyama, C.; Wu, J.; Briggs, M.R.; Brown, M.S.; Goldstein, J.L.; Wang, X. Proc. Natl. Acad. Sci. U.S.A. (1993) 90:11603-11607

#journal SREBP-2, a second basic-helix-loop-helix-leucine zipper protein that stimulates transcription by binding to a sterol regulatory element.

#title

#cross-references MUID:94085681

#accession A49397

#molecule_type mRNA

#residues 1-1141 ##label HUA

##cross-references GB:U02031; NID:9451329; PIDN:AA50746.1; PID:9451330

##experimental_source HeLa cells

##note Sequence extracted from NCBI backbone (NCBIP:140987)

REFERENCE A54962

#authors Yang, J.; Sato, R.; Goldstein, J.L.; Brown, M.S. Genes Dev. (1994) 8:1910-1919

#journal Sterol-resistant transcription in CHO cells caused by gene rearrangement that truncates SREBP-2.

#title

#cross-references MUID:95047343

#contents annotation

COMMENT Under conditions of sterol depletion, this membrane-bound transcription factor is released from extranuclear membranes to activate sterol biosynthesis.

GENETICS

#gene GDB:SREBF2; SREBP2

##cross-references GDB:273625

#map_position Xq24-Xqter

CLASSIFICATION #superfamily sterol regulatory element binding protein

KEYWORDS DNA binding; membrane protein; transcription factor

SUMMARY #length 1141 #molecular_weight 123672 #checksum 4172

Query Match 15.6%; Score 1328; DB 2; Length 1141;

Best Local Similarity 30.2%; Pred. No. 2,41e-189;

Matches 320; Conservative 270; Mismatches 405; Indels 66; Gaps 50;

Db 118 PTSVPTPRATPILOPRPOPOPOLOQOQVMI-TP-TFSTTPORRIIOQPL-IVONA 174

QY 74 POSVEQOPHYKSEHSPVHIKEELHQQOQSPLLVYKPDPLIATSYNCPQOQPTGLKAA 133

Db 175 ATRSOVLQPOVQSLVTSQOQPVITIQOQ-V-QTVQAOQVNLQTNQNGITQTLAPAVQVVA 232

QY 134 QPTATIHMDAQRPPTAVYPPSLGSSFYQSSPSPSPESANQNVNMQP--VAATP 191

Db 233 APQVQVPLVQPOIIRKIDSLVLTLLKTDGSPVMAVONPALATPTPTQTAALQVPTLV 292

QY 192 APAAPLPQOSYPOFITYN-SKAGMSDEAMYL-LO-PIVAS-PTSPSPVA-PPPIST 246

Db 293 GSGSTIITPMVMGQEKVPIKQVPGVYKOLEPPEKGGRTTHNIIEKRYSSINDKITE 352

QY 247 GSRASKYRVAPLAPSPAMEV-QCKVPIINVOQPKYKVKSAHNAIERRTSINDKINE 305

Db 353 LKDLVMTDAKMHKSGYLARKLIDYIKYQOVNHLROENMVL-K-LANQKKLLKGI-DL 409

QY 306 LKNLVGEQALNKSVAVRIKSIDIRDLRONHDLKALQRLQRLAMRDGSKYVDLQL 365

Db 410 GSLVDNEVDLKE--D-FNQVLLMSPASDSG-SQAGF-SPYSIDSEP--GSPILL--D 459

QY 366 GTRPGASKKRRRESSQFTTDTAGLTPRSDSDSLSPMHDSILPSPYGGSTASCSSG 425

Db 460 DAKYKDEPSPVAL-GWVDSRIILCVLPFCSPNPLTSLQMGAGHSDQPHSGG 518

QY 426 SSSNEEPLVPPSSMRGNATHSRGLCMFMFAILLAVNPFKFLQGRHDSNDLDGMSGQ 485

Db 519 RSVLSFESGGGWFDMMPITLLMLVNGVIVLSVFKLLVGEPEVIRPHSRSSVTFMHR 578

QY 486 RRIISYDVECEGFVAVMOSSMI-WLNFPTMLGLCVLLVYGDQL--DAQTD-AYCOHR 541

Db 579 KQADIDLARCFPAANAANLOTCLAVLGRALPTSLDLACSL--SMNVIKISLQRLVW 636

QY 542 QRADFYSGQSGQAYGYLNCIMFGLSLPASRLF--CYLQTTWQFLRFHFH--RL--W 595

Db 637 LKKVFOCRIRATPATPEGEFDEAKTSADALAVHRIHQHITGKLPAGSACSVYHMLC 696

QY 596 L-GVLSRRSGGLFSNASKQALASARELALLENRLQLOLTGNGSGDN-NGIMMALF 653

Db 697 AVNLAECAEEKIPSPSTVEIHILTPAMGILTRCGK-Ig-FLASYFLSRAQSLCGPERSAV 754

QY 654 ASMAEVAHNLITREFICIHVMALRMK-RSARKWLDQFFARYMSARQDCRTATE 712

Db 755 P-DSLRLWICHPLGQKFMER--SWSVK-SAAKESLYCAQRNPADPIAQVHOAFCKMLER 810

QY 713 QTOELRMWAFAYGYRCATVHFYDLSDSGBDGFFTRLRNPCDPAHVIKOYREHLFLK 772

Db 811 AIESLVNQAQAKKKGDEEESCEFSALAEYIKLHSPVDSGVNVPSPLSRSVYLSALGP 870

QY 773 SIQCLVGA-GHKSGGLPTSSVSGAEQLOQOOSHGTIVSNV-LVYTSILKLTQWADER 830

Db 871 DIORWMTSATITVAISWLOGDAVRSHTFVERIPALF-VTES-PLVKAIFPAC-RAM 927

QY 831 DTNVMWADVETAVHMLIGEDTLAEQLYGRIKMPLOLOGCGENHPLKL-HAVLRK 889

Db 928 HASLPGRADG-QQS-S-FCH-CERASGHLMSLVNSGGTSPALNHVYLLTCDLLSL 982

QY 890 MILKKNQNALDKSLQVNLICDESSVEIQELTVNRITDAKIKILFQLTCDMLLET 949

Db 983 RIALMQOASASQAVGEFYHNASGAEIAGFORDGLSLRLHSPAPAKVYLHATVRLM 1042

QY 950 RTALWELHNMEDDG-FYQVPEVLEKFFQDLMSLRIVENIINASRIYLYAVCYRLM 1008

Db 1043 AGSPTRTHOLESLRTRTQSTKH-GEVDAMP-----GGERATATILACRHLPSFL 1096

QY 1009 AGASPCPTQLDLSLSRRAHSSIFCGSKDRROONTVGGGERASAMVACKYLPALL 1068

Db 1097 SSPQRAVLAAEARLLEKVGDRRSCNDQOQIVKLGGA 1137

QY 1069 SSPGERAGMLAEAKTLEKVGDRKRLKECYQLKMSLNGIG 1109

RESULT 4

ENTRY B54962 #type complete

TITLE sterol regulatory element binding protein 2 precursor - Chinese hamster

ORGANISM #formal_name Cricetulus griseus #common_name Chinese hamster

DATE 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 04-Sep-1998

ACCESSIONS B54962

REFERENCE A54962

#authors Yang, J.; Sato, R.; Goldstein, J.L.; Brown, M.S. Genes Dev. (1994) 8:1910-1919

#journal Sterol-resistant transcription in CHO cells caused by gene rearrangement that truncates SREBP-2.

#title

#cross-references MUID:95047343

#accession B54962

##status preliminary

##molecule_type mRNA

##residues 1-1139 ##label YAN

##cross-references GB:U12330

##note 493 Ser was also found

CLASSIFICATION #superfamily sterol regulatory element binding protein

Db	735	RF-EVVDG-DMAV-ICAPDESLYSVAGNVPDLAOVTRLFCEHLEERLNCIAOPSPDAD	791
Oy	727	RYCATHTVTTYLSOSGEDDGFITLRNCPDAAHYIORYRHLLFKSTIOCLVAGHSGG	786
Db	792	-GHRFSDALGIOLNLNSCDVAGAPA-CSFSVSSM-ATTGTDPYAKWASLTAVYIH	848
Oy	787	LPTSSVSGAEOLQOOSCTIVNLEKYTLTKDTLMADEDDEDTNVMMADVLETAHV	846
Db	849	WLRDEEAERLPYLVHIIPOVDO	873
Oy	847	WLGEDTLAEOLYGRIKOMPTOLOO	871
RESULT	5		
ENTRY		PD0035	#type complete
TITLE			sterol regulatory element-binding protein-1 isoform - mouse
ORGANISM			#formal_name Mus musculus #common_name house mouse
DATE			23-Jul-1999 #sequence_revision 23-Jul-1999 #text_Change 20-Aug-1999
ACCESSIONS		PD0035	
REFERENCE		PD0035	
authors			Inoue, J.; Sato, R.
journal			Biosci. Biotechnol. Biochem. (1999) 63:243-245
title			A novel splicing isoform of mouse sterol regulatory element-binding protein-1 (SREBP-1).
#cross-references		M01D:99161303	
accession		PD0035	
status			Preliminary
##molecule_type		mRNA	
##residues		1-391	#label INO
##cross-references		DDAJ:AB017337; NID:g4240011	
transcription			
length		391	#molecular_weight 39912 #checksum 4894
SUMMARY			

[illegible]

		stabilize both DNA binding and multimerization.	
#cross-references	MUID:92123207		
#accession	A42029		
#status	preliminary		
#molecule_type	mRNA		
##residues	1-446 ##label ROM		
##cross-references	GB:S76673; NID:q243439; PID:q243440		
#note	sequence extracted from NCBI backbone (NCBIN:76673, NCBIPI:76674)		
SUMMARY	#length 446 #molecular-weight 47891 #checksum 8873		
Query Match	2.4%; Score 192; DB 2; Length 446;		
Best Local Similarity	25.7%; Pred. No.1,01e-08;		
Matches	64; Conservative 65; Mismatches 100; Indels 20; Gaps 16;		
Db	42 PAQVPREVLKAYTHLENPRTYHLQOARRQVKQYLTSLGPKLASQALPPPGSSAQPL 101		
QY	122 PQOQOPTGLTKA-AQ-PTATIIHMDAQMPNPNTAYVPSLTGSSFYQMSPTSP--VES- 176		
Db	102 PAPFAHAHTGTSAGAPRSPMALLTGSSSEKEDVIDELIISLSSYNDMLSTLPQGT 161		
QY	177 -ANQNVNMMQVAAATPAPASAPLPQOOSYPOPEI--IYNSAGM-TS-DEAMYLLOPTVA 221		
Db	162 GLQLPSTLPVSGNLLDVYSSQGVATPAITYNSQCPALPIKREISETFEKALKEROK 221		
QY	232 S-PTPPSP-PVAPPTST-GSRASVVRRAPLAPS-PALM-EVQGVNPINRQPKAYEV-KR 285		
Db	222 DNHNLIERRRRFINDRIKELGITLIPKSNDEPMKWKGTILKASVDYIRKLQEQQRSK- 280		
QY	286 SAHAHAIERRYTSTINDKINELKNLVV-GEQAKL--NKSAVLRKSIDKIRDLQRONHDLKA 342		
Db	281 DLESRRSL 289		
QY	343 ELQRLQREL 351		

RESULT	8	
ENTRY		138024
TITLE		MTF protein - human
ORGANISM		#formal_name Homo sapiens #common_name man
DATE		17-May-1996 #sequence_revision 17-May-1996 #text_change 29-Aug-1997
ACCESSIONS		138024
REFERENCE		138024
authors		Tachibana, M.; Perez-Jurado, L.A.; Nakayama, A.; Hodgkinson, C.A.; Li, X.; Schneider, M.; Miki, T.; Fex, J.; Francke, U.; Arnheiter, H.
#journal		Hum. Mol. Genet. (1994) 3:553-557
#title		Cloning of MTF, the human homolog of the mouse microphthalmia gene and assignment to chromosome 3p14.1-1p12.3.
#cross-references		MTD:94348499
#accession		138024
#status		preliminary: translated from GB/EMBL/DBJ
#molecule_type		mRNA
#residues		1-419 #label RES
#cross-references		EMBL:Z29678; NID:g468496; PID:g468497
GENETICS		
#gene		GDB:MTF
#cross-references		GDB:214776; OMIM:156845; OMIM:193510
#map_position		3p14.1-3p12
SUMMARY		#length 419 #molecular_weight 46938 #checksum 307
Query Match		2.3%; Score 182; DB 2; Length 419;
Best Local Similarity		40.0%; Pred. No. 1,81e-07;
Matches		32; Conservative 22; Mismatches 21; Indels 5; Gaps 4;
Db	197	RALAKERO-KKDNILRRRRRINDRIKELGTLIPKSNPDMRWKGTILKASVDYIR 255
Oy	275	RVQPKVKEVKSANNAIEPRYRSTINDKINELKLVV-GEQATL--NKSAYLKRSLDKIR 331
Db	256	KLQREQQRAR-ELLENRKKL 274

```

OY      332 DLORONHDLKAELOREL 351

RESULT      9
ENTRY      A40728 #type complete
TITLE      microphthalmia-associated transcription factor ml - mouse
ORGANISM   #formal_name Mus musculus #common_name house mouse
DATE       19-May-1994 #sequence_revision 19-May-1994 #text_change
          10-Sep-1997

ACCESSIONS A40728; A48021; I49245; S36645
REFERENCE   A40728
#authors    Hodgkinson, C.A.; Moore, K.J.; Nakayama, A.; Steingrimsson,
#journal     E.; Copeland, N.G.; Jenkins, N.A.; Arnheiter, H.
#title       Cell (1993) 74:395-404
#title       Mutations at the mouse microphthalmia locus are associated
#title       with defects in a gene encoding a novel
#title       basic-helix-loop-helix zipper protein.
#cross-references MUID:93345026
#accession   A40728
#status      preliminary
#molecule_type mRNA
#residues    1-419 ##label HOD
#cross-references GB:23066; NID:9396739; PID:9396740
REFERENCE   A48021
#authors     Hughes, M.J.; Lingrel, J.B.; Krakowsky, J.M.; Anderson, K.P.
#journal     J. Biol. Chem. (1993) 268:20687-20690
#title       A helix-loop-helix transcription factor-like gene is located
#title       at the ml locus.
#cross-references MUID:94012591
#accession   A48021
#status      preliminary
#molecule_type DNA
#residues    238-286 ##label HUG
#cross-references GB:I22958
REFERENCE   I49244
#authors     Steingrimsson, E.; Moore, K.J.; Lamoreux, M.L.;
#journal     Ferre-D'Amaré, A.R.; Burley, S.K.; Sanders Zimring, D.C.;
#title       Skow, L.C.; Hodgkinson, C.A.; Arnheiter, H.; Copeland,
#title       N.G.; Jenkins, N.A.
#title       Nature Genet. (1994) 8:256-263
#title       Molecular basis of mouse microphthalmia (ml) mutations helps
#title       explain their developmental and phenotypic consequences.
#cross-references MUID:95179171
#accession   I49245
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues    1-16 ##label RES
#cross-references EMBL:U19875; NID:9642572; PID:9642573
GENETICS
#gene        ml
#keywords     DNA binding; transcription factor
SUMMARY      #length 419 #molecular-weight 46768 #checksum 1668

Query Match      2.3%; Score 181; DB 2; Length 419;
Best local Similarity 38.8%; Pred. No. 2.41e-07;
Matches 31; Conservative 23; Mismatches 21; Indels 5; Gaps 4;

DB      197 BALAKERO-KKDNHNLIRRRRNINDRIKELGTLIPKSNPDPMRNKGTILKASVDYIR 255
      1 : : : : : | | | | | : | | | | | : : : : : | | : | | |
OY      275 RVQPKYKEVKRSANHAIERRYRTSINDKINELKNLV-GEQAKL--NKSAVLRKSIDKIR 331
      | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

DB      256 KLOREOQRAK-DLENROKKL 274
      | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY      332 DLORONHDLKAELOREL 351
      | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT      10
ENTRY      T14752 #type complete
TITLE      microphthalmia-associated transcription factor, MITF-A -
ORGANISM   human
#formal_name Homo sapiens #common_name man
#sequence_revision 20-Sep-1999 #text_change
DATE       24-Sep-1999

ACCESSIONS T14752; JE0207
REFERENCE   T14752
#authors    Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.;
#journal     Wiemann, S.
#title       Submitted to the Protein Sequence Database, August 1999
#accession   T14752
#status      preliminary
#molecule_type mRNA
#residues    1-520 ##label KOE
#cross-references EMBL:AL110195
#experimental_source adult uterus; clone DKFZp586B2217
REFERENCE   JE0207
#authors     Amae, S.; Fuse, N.; Yasumoto, K.; Sato, S.; Yajima, I.;
#journal     Yamamoto, H.; Udono, T.K.; Durlu, Y.; Tamai, M.; Takahashi,
#title       K.; Shibahara, S.
#title       Biochem. Biophys. Res. Commun. (1998) 247:710-715
#title       Identification of a novel isoform of
#title       microphthalmia-associated transcription factor that is
#title       enriched in retinal pigment epithelium.
#cross-references MUID:98321192
#accession   JE0207
#molecule_type mRNA
#residues    1-520 ##label AMA
#cross-references DDBJ:AB006909
COMMENT     This protein plays a important role in differentiation of retinal
            pigment epithelium and regulates melanogenesis in retinal pigment
            epithelium.
GENETICS
#note        DKFZp586B2217.1
SUMMARY      #length 520 #molecular-weight 58162 #checksum 6813

Query Match      2.3%; Score 182; DB 2; Length 520;
Best local Similarity 40.0%; Pred. No. 1.81e-07;
Matches 32; Conservative 22; Mismatches 21; Indels 5; Gaps 4;

DB      298 BALAKERO-KKDNHNLIRRRRNINDRIKELGTLIPKSNPDPMRNKGTILKASVDYIR 356
      1 : : : : : | | | | | : | | | | | : : : : : | | : | | |
OY      275 RVQPKYKEVKRSANHAIERRYRTSINDKINELKNLV-GEQAKL--NKSAVLRKSIDKIR 331
      | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

DB      357 KLOREOQRAK-ELENROKKL 375
      | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY      332 DLORONHDLKAELOREL 351
      | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT      11
ENTRY      A34596 #type fragment
TITLE      transcription factor E3 - human (fragment)
ORGANISM   #formal_name Homo sapiens #common_name man
DATE       06-Jul-1990 #sequence_revision 09-Oct-1992 #text_change
          24-Sep-1998
ACCESSIONS A34596; S10379
REFERENCE   A34596
#authors     Beckmann, H.; Su, L.K.; Kadesch, T.
#journal     Genes Dev. (1990) 4:167-179
#title       TFE3, a helix-loop-helix protein that activates transcription
#title       through the immunoglobulin enhancer mub3 motif.
#cross-references MUID:90249724
#accession   A34596
#status      preliminary
#molecule_type mRNA
#residues    1-536 ##label BEC
#cross-references EMBL:X51330; NID:937061; PID:e20907; PID:g1335343
#note        the authors translated the codon ACC for residue 433 as
            Ser, and GAG for residue 472 as Gln

GENETICS
#gene        GDB:TFE3
#cross-references GDB:125870; OMIM:314310
#map_position XP11.23-XP11.22
KEYWORDS     DNA binding; transcription factor
SUMMARY      #length 536 #checksum 8782

Query Match      2.1%; Score 172; DB 2; Length 536;
Best local Similarity 40.8%; Pred. No. 3.06e-06;
```

Matches	29;	Conservative	19;	Mismatches	19;	Indels	4;	Gaps	3
Db	140	KKDHNHLEIRRRRNINRIKELGTLIPKSSDPEWRNKGITLKSADYIRLQKQDORS	199						
Oy	284	KRSANMAIERYRFSINDIKNELNLV--VGE-QAKLNKSAVLKRSIDKIRDLORNDL	340						
Db	200	K-DLESRQSL	209						
Oy	341	KAELOFLOREL	351						
RESULT	12								
ENTRY	A35558	#type fragment							
TITLE	transcription factor TFEB - human (fragment)								
ORGANISM	#formal_name Homo sapiens #common_name man								
DATE	28-Sep-1990 #sequence_revision 28-Sep-1990 #text_change 16-Feb-1997								
ACCESSION	A35558								
REFERENCE	A35558								
#authors	Carr, C.S.; Sharp, P.A.								
#journal	Mol. Cell. Biol. (1990) 10:4384-4388								
#title	A helix-loop-helix protein related to the immunoglobulin E box-binding proteins.								
#cross-references	MUID:90318407								
#accession	A35558								
#status	preliminary								
#molecule_type	mRNA								
#residues	1-514 #label CAR								
#cross-references	GB:M3782								
KEYWORDS	DNA binding; transcription regulation								
SUMMARY	#length 514 #checksum 4657								
Query Match	2.0%;	Score 162;	DB 2;	Length 514;					
Best Local Similarity	23.7%;	Pred. No. 4,83e-05;							
Matches	71;	Conservative	88;	Mismatches 115;	Indels	26;	Gaps	22;	
Db	111	QAQGEQRERMQQAAMHYMQQ-QQQQQQQQLG-PPPAINTPVRHFQSPVPYGEVLKV	168						
Oy	75	QSVEQPVRKSEHSPVRIKEELHQ0000SPFLVYKPPDLIAT--YNCPOQOPTGLKA	132						
Db	169	QSLLENPTSYHLOQSQHKVREYLSFTYGNKFFAAHISPAQALBNPQHPQGGELDTCC	228						
Oy	133	-A-Q-PTA-TIHMDQORMPN-TAVYPSLGSFFY--QSM-SPTSPESANQWNV	183						
Db	229	PPPLATVLPSPKMLHIGSNPERELDDVINIRLIMSIGYINPEQMNTLPLSSSL	288						
Oy	184	MQPVAAAT-PARASAPLQOQSYPQFFITYSAGMTSDEAM-YL-L-LQPTVASPTPEPV	239						
Db	289	NVYSSDPVATSLGVYTS-SSCPADLTQKRELITDAESRALAKERO-KKDHNHLEIRRRF	346						
Oy	240	APPPTSTGRSASKYRVAFLAPSPAA--EVQG-KVPIRWQPKYEVKRSANNAIERYRT	297						
Db	347	NINRIKELGMLIKANDLDVRNKGITLKSADYIRRMQDKQSR-ELENHSRLQEMT	405						
Oy	298	SINDIKNELNLV--VGE-QAKLNKSAVLKRSIDKIRDLORNDLKAELQLOREL-MA	353						
RESULT	13								
ENTRY	T04073	#type complete							
TITLE	intensifier 1 protein - maize								
ORGANISM	#formal_name Zea mays #common_name maize								
DATE	23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Apr-1999								
ACCESSION	T04073								
REFERENCE	215188								
#authors	Burr, F.A.; Burr, B.; Scheffler, B.E.; Blewitt, M.; Wienand,								
#journal	U.; Matz, E.C.								
#title	Plant Cell (1996) 8:1249-1259								
#accession	T04073								
#status	preliminary; translated from GB/EMBL/DBJ								

	#molecule_type DNA	
	##residues 1-685 ##label BUR	
	##cross-references EMBL:057899; NID:g1420923; PID:g1420924	
	##experimental_source strain W22	
GENETICS		
#gene	in1	
#introns	36/7: 148/2; 180/3; 185/3; 204/3; 343/1; 496/3; 650/3	
SUMMARY	#length 685 #molecular_weight 74698 #checksum 2432	
Query Match	2.0%; Score 159; DB 2; Length 685;	
Best Local Similarity	22.9%; Pred. No. 1.09e+04;	
Matches	36; Conservative 51; Mismatches 62; Indels 8; Gaps 7;	
Db	423 APSSSHSIHNGEVOSSPEPRGDDGGCTSRKRGVPVPSQTELS-AshVLEKKRRREKLN 481	
Oy	242 PPTSGSASAKSVRAPIAPSPAAAEVGKVINRVQ-PKYKEVKRSAHNAIERRYRTISN 300	
Db	482 EGFAFLRSLFVFY-TKMDRASILDTIEYKOLRRRIQLQ-E-SR-RRLVGSNOKTTM 535	
Oy	301 DKINELKLVVEADAKLKSAVLKRSDIKRDJROKHMDKAELQRLQRELMAKDSGYK 360	
Db	536 AQQPPPAASTERGERROTSGGYLARAGTGRSAEA 572	
Oy	361 DLGLGTGRGRASKRKRESSOTFTTDGLLPQRDES 357	
RESULT 14	A42611 #type complete	
ENTRY TITLE	basic-helix-loop-helix-leucine zipper protein delta Max - human	
ALTERNATE_NAMES	max protein 2	
ORGANISM	#formal_name Homo sapiens #common_name man	
DATE	04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 07-May-1999	
ACCESSIONS	A42611; S33119	
REFERENCE	A42611 Mañela, T.P.; Koskinen, P.J.; Vastrik, I.; Allitalo, K.	
#authors	Science (1992) 256:373-377	
#journal	Alternative forms of Max as enhancers or suppressors of MYC-ras cotransformation.	
#title	Nucleic Acids Res. 20(1992) 1001-1006	
#cross-references	MIMD:92229468	
#accession	A42611	
##status	Preliminary	
##molecule_type	mRNA	
##residues	1-103 ##label MAX	
##cross-references	GB:X60287; GB:S95058; NID:q599792; PID:q599793	
##experimental_source	Erythroblastic leukemia cells HEL	
##note	"Sequence extracted from NCBI backbone (NCBIN:95058, NCBI:P:95059)"	
REFERENCE	S33118	
#authors	Vastrik, I.; Koskinen, P.J.; Allitalo, R.; Maekelae, T.P.	
#journal	Oncogene (1993) 8:503-507	
#title	Alternative mRNA forms and open reading frames of the max gene.	
#cross-references	MIMD:93149613	
#accession	S33119	
##status	Preliminary	
##molecule_type	DNA	
##residues	1-103 ##label VAE	
##cross-references	EMBL:X6867; NID:q34469; PID:q34471	
GENETICS		
#introns	99/1	
KEYWORDS	alternative splicing	
SUMMARY	#length 103 #molecular_weight 12099 #checksum 6857	
Query Match	1.9%; Score 149; DB 2; Length 103;	
Best Local Similarity	29.9%; Pred. No. 1.55e+03;	
Matches	26; Conservative 28; Mismatches 30; Indels 3; Gaps 3;	
Db	7 IEVSDEPQPFPOSA-D-KRAHNALERKRDPHIKSPLSDSVPSLOGEKASPQAII 64	
Oy	265 MEVQGKPVRNVQPKVKEKRSANAHERYRTISINDKINELNLVAIGDA-KLNKSAVL 323	

Db 65 DKATEYIOYMRKKNHTHOODIDDLKRQ 91
 QY 324 RKSIDKIRDLOQONHDLKAELOQLQRE 350

RESULT 15

ENTRY S33120 #type complete
 TITLE max protein 3 - human
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-May-1999

ACCESSIONS

S33120
 S33118

REFERENCE

#authors Vaestriik, I.; Koskinen, P.J.; Allitalo, R.; Maekelae, T.P.
 #journal Oncogene (1993) 8:503-507
 #title Alternative mRNA forms and open reading frames of the max gene.

#cross-references MUID:93149613
 #accession S33120

#status preliminary

#molecule-type DNA

#residues 1-134 #label VAE

##cross-references EMBL:X66867; NID:934469; PID:934472

SUMMARY #length 134 #molecular-weight 15395 #checksum 5928

Query Match 1.9%; Score 149; DB 2; Length 134;

Best Local Similarity 29.9%; Pred. No. 1.55e-03;

Matches 26; Conservative 28; Mismatches 30; Indels 3; Gaps 3;

Db 7 TEVESDEEQPRFOSAA-D-KRAHNALERKRDRDHKDSFHSLSRDSVPSLOGEKAQRAQIL 64

QY 265 MEVOGKVPINRYOPKVEKNSAHNAIERRYRTSINDKINELKNLVGEQA-KLNKSAVL 323

Db 65 DKATEYIOYMRKKNHTHOODIDDLKRQ 91

QY 324 RKSIDKIRDLOQONHDLKAELOQLQRE 350

Search completed: Sun Sep 3 12:52:55 2000
 Job time : 112 secs.


```

Db 241 PPPTSTGRASRVAPLAPSPAMEVOGKVPINRQPKVKEKRSAAHATERRTSIN 300
QY 241 PPPTSTGRASRVAPLAPSPAMEVOGKVPINRQPKVKEKRSAAHATERRTSIN 300
Db 301 DKINELKMLVYGEQAKLNKSAVLKRSIDKIRDLQONHDLKAELORLQRELMARGSKYK 360
QY 301 DKINELKMLVYGEQAKLNKSAVLKRSIDKIRDLQONHDLKAELORLQRELMARGSKYK 360
Db 361 DLLOGTGPRGSKRRRESSOTFTDAGLTPRPSDESPTSIPMSHSDISLPPSPGGSTA 420
QY 361 DLLOGTGPRGSKRRRESSOTFTDAGLTPRPSDESPTSIPMSHSDISLPPSPGGSTA 420
Db 421 SCSSGSSSSNEEPLVYVPSMRGMATHSRGLGCMFALVAVPFTFLORGHYSDNDLG 480
QY 421 SCSSGSSSSNEEPLVYVPSMRGMATHSRGLGCMFALVAVPFTFLORGHYSDNDLG 480
Db 481 DMSGORLLSDVDEGEGRVAVWOSSWMLNTLMLGCLVLLVYGDPLDQADTAYCOH 540
QY 481 DMSGORLLSDVDEGEGRVAVWOSSWMLNTLMLGCLVLLVYGDPLDQADTAYCOH 540
Db 541 RORADFYRSOGSSOAYVGYLNCILMFGLSPASRLECYLQTTWQFLRFLRLMGLV 600
QY 541 RORADFYRSOGSSOAYVGYLNCILMFGLSPASRLECYLQTTWQFLRFLRLMGLV 600
Db 601 SRSRSGLSFNMAASRQALASARELALLFNRLNQLDLTGNSRGDNGIMMALFASNMAEV 660
QY 601 SRSRSGLSFNMAASRQALASARELALLFNRLNQLDLTGNSRGDNGIMMALFASNMAEV 660
Db 661 AANLITPPEITICHYMTALRKRSAPKWLQOPFARYMSRAOEGGRTRATQOTELRNA 720
QY 661 AANLITPPEITICHYMTALRKRSAPKWLQOPFARYMSRAOEGGRTRATQOTELRNA 720
Db 721 FTAYGYRCATHEFTYDLSDSGEGDGFTRLNPCDPAHVTKOYREHLFFSTOCIVA 780
QY 721 FTAYGYRCATHEFTYDLSDSGEGDGFTRLNPCDPAHVTKOYREHLFFSTOCIVA 780
Db 781 GHKSGGLTSSVSGEAEOLQOQOHSQTIYSNLYKTSILKDLTAMDDEDERDNTVVMADY 840
QY 781 GHKSGGLTSSVSGEAEOLQOQOHSQTIYSNLYKTSILKDLTAMDDEDERDNTVVMADY 840
Db 841 LETAVHMLGEBDTLAEOLYGRKQMPLOOQCGENDHLPKALHAVLRAKMITLKNNGAL 900
QY 841 LETAVHMLGEBDTLAEOLYGRKQMPLOOQCGENDHLPKALHAVLRAKMITLKNNGAL 900
Db 901 DKSJLQVNIILCDESSVELQECTVNRITDAKIKLFCOLLCDWLLETRTALWLEHYN 960
QY 901 DKSJLQVNIILCDESSVELQECTVNRITDAKIKLFCOLLCDWLLETRTALWLEHYN 960
Db 961 MEDDFYQVPGFVLEKFPQDLSLNLIENIENIPNAOSRIYLYEAVCRMLAGASPCFTQOL 1020
QY 961 MEDDFYQVPGFVLEKFPQDLSLNLIENIENIPNAOSRIYLYEAVCRMLAGASPCFTQOL 1020
Db 1021 DRSJLSRNAHSSIFCGSDRROQNVGGEREPASAMYVACKTLPALLSSPGERAGMLAE 1080
QY 1021 DRSJLSRNAHSSIFCGSDRROQNVGGEREPASAMYVACKTLPALLSSPGERAGMLAE 1080
Db 1081 AAKTLEKVGDKRKLECYQLMKSLGNGISVYA 1113
QY 1081 AAKTLEKVGDKRKLECYQLMKSLGNGISVYA 1113

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99161303.
RA INOUE J., SATO R.;
RT "A novel splicing isoform of mouse sterol regulatory element-binding
protein-1 (SREBP-1).";
RL Biosci. Biotechnol. Biochem. 63:243-245(1999).
DR EMBL: AB017337; BAA74795.1; -.
FT NON TER 1
FT 403
SO SEQUENCE 403 AA; 41018 MW; BB57AA8F CRC32;

Query Match
Best Local Similarity 34.8%; Pred. No. 6,11e-33;
Matches 85; Conservative 58; Mismatches 86; Indels 15; Gaps 12;

Db 153 PTPALHTOVQSLASQQLPAPSAAPRTNVTISQVOQVPLVLOPHRTKADSLTAVKVDAG 212
QY 170 PTPVESANQVNWQVPAATPAPASAPL-PO-QSYP--Qp-FITYNSKA-GMTSDAM 222
Db 213 ATVTAGISTLAPGTAVAGPLQTLVSGTILATVPLVVDTKLPIHRLAAGSKALGSAQ 272
QY 223 YLLQPTVASTPTSPVPVAPPTSTGSRASKVRA-PLAPSPAMEVOGKVPINR-V-QPK 279
Db 273 SREKRTAHNAIEKRRSSINDKIVEKLDVYGTAKLNKSAVLKALDITRFLQHSNOK 332
QY 280 VKEYRSAHNAIERRYRTSINDKINELKMLVYGEQAKLNKSAVLKRSIDKIRDLQRONHD 339
Db 333 LKQENTLT-RS-AHKSLSLDVYS-ACGSGGCTDVSNECKKPEVER-LTPPSDAGSP 387
QY 340 LKAELORLQRELMARGSKVADLLQLTGPRGSKRRRESSOTFTDAGLTPPSDESDEP 399
Db 388 SSSS 391
QY 400 SLEP 403

RESULT 3
ID 09XX00 PRELIMINARY: PRT: 1113 AA.
AC 09XX00:
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
DE Y47D3B.7 PROTEIN.
GN Y47D3B.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA MATTHEWS L.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAYELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIKKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SKALDON N., SMITH A., SONNHAMMER E., STRADER R., SULLSTON J.,
RA THIERER-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLIDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
DR EMBL: AL031635; CAA21042.1; -.
SO SEQUENCE 1113 AA; 125156 MW; BA56E60 CRC32;

```


RN RP SEQUENCE FROM N.A.
 RX MEDLINE: 98167856.
 RA HODGKINSON C.A., NAKAYAMA A., LI H., SWENSON L.-B., OPDECAMP K.,
 RT ASHER J.A. JR., ARNHEITER H., GLASER T.;
 RT "Mutation at the anophthalmic white locus in Syrian hamsters:
 RT haploinsufficiency in the Mltf gene mimics human Waardenburg syndrome
 RT type 2.";
 RL Hum. Mol. Genet. 7:703-708(1998).
 DR EMBL: AF020900; AAC15952.1; -.
 DR HSSP: P22415; 1AN4.
 DR PFAM: PF00010; HLH; 1.
 FT NON_TER 1
 SQ SEQUENCE 377 AA; 41728 MW; 83722925 CRC32;
 Query Match 2.3%; Score 181; DB 11; Length 377;
 Best Local Similarity 38.8%; Pred. No. 1,03e-08;
 Matches 31; Conservative 23; Mismatches 21; Indels 5; Gaps 4;
 Db 155 BALAKERO-KKDNNHLIERRRRNINDRIKELGTLIPKSNDDPMRWKGTILKASVDYIR 213
 I : : : : : | | | | | : | | | | | : : : : : | | | | |
 Oy 275 RVQPPVKVKKSAHNAIERRYRTSINDKINLKNLVV-GEQAKL--KKSATLKRSIDKIR 331
 I : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
 Db 214 KLQEOQRAK-DLENROKKL 232
 I : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
 Oy 332 DLORONHDLKAELOREL 351
 I : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
 RESULT 7
 ID 014841 PRELIMINARY; PRT; 419 AA.
 AC 014841;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
 DE MTF PROTEIN (MICROPHthalmIA-ASSOCIATED TRANSCRIPTION FACTOR).
 GN MTF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE-SKIN;
 RX MEDLINE: 94348499.
 RA TACHIBANA M., PEREZ-JURADO L.A., NAKAYAMA A., HODGKINSON C.A., LI X.,
 RA SCHNEIDER M., MIKI T., FEX J., FRANCKE U., ARNHEITER H.;
 RT "Cloning of MTF, the human homolog of the mouse microphthalmia gene
 RT and assignment to chromosome 3p14.1-p12.3.";
 RL Hum. Mol. Genet. 3:553-557(1994).
 RN [2]
 RP SEQUENCE OF 1-11 FROM N.A.
 RX MEDLINE: 98160190.
 RA WATANABE A., TAKEIDA K., PLOPLIS B., TACHIBANA M.;
 RT "Epistatic relationship between Waardenburg syndrome genes MTF and
 RT PAX3.";
 RL Nat. Genet. 18:283-286(1998).
 DR EMBL: Z29678; CA82775.1; -.
 DR EMBL: AF034755; AAC39639.1; -.
 DR HSSP: P22415; 1AN4.
 DR PFAM: PF00010; HLH; 1.
 KW DNA-binding.
 SQ SEQUENCE 419 AA; 46938 MW; 5CE200FD CRC32;
 Query Match 2.3%; Score 182; DB 4; Length 419;
 Best Local Similarity 40.0%; Pred. No. 7,48e-09;
 Matches 32; Conservative 22; Mismatches 21; Indels 5; Gaps 4;
 Db 197 BALAKERO-KKDNNHLIERRRRNINDRIKELGTLIPKSNDDPMRWKGTILKASVDYIR 255
 I : : : : : | | | | | : | | | | | : : : : : | | | | |
 Oy 275 RVQPPVKVKKSAHNAIERRYRTSINDKINLKNLVV-GEQAKL--KKSATLKRSIDKIR 331
 I : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
 Db 256 KLQEOQRAK-ELENROKKL 274

[illegible]

RC STRAIN-C57BL/6; TISSUE=HEART;
 RX MEDLINE; 9519171.
 RA STEINERINSON E., MOORE K.J., LAMOREUX M.L., FERRE-D'AMARE A.R.,
 RA BURLEY S.K., SANDERS ZIMRING D.C., SKOW L.C., HODKINSON C.A.,
 RA ARNHEITER H., COPELAND N.G., JENKINS N.A.; mutations helps explain
 RT "Molecular basis of mouse microphthalmia (mi) mutations helps explain
 RT their developmental and phenotypic consequences."
 RL Nat. Genet. 8:256-263(1994).
 CC -1- FUNCTION: ACTS AS A TRANSCRIPTION FACTOR.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH TFE3 OR TFE8 (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN, ISOFORMS 1 AND 2,
 CC ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE. THE
 CC SEQUENCE SHOWN HERE IS THAT OF ISOFORM 1.
 CC -1- TISSUE SPECIFICITY: IN THE ADULT, EXPRESSED AT HIGH LEVELS IN THE
 CC HEART, SKIN AND SKELETAL MUSCLE. BARELY DETECTABLE IN OTHER
 CC TISSUES. IN THE EMBRYO, EXPRESSED IN DEVELOPING EYE, EAR, SKIN AND
 CC HEART.
 CC -1- DISEASE: DEFECTS IN THE MITF GENE ARE THE CAUSE OF MICROPTHALMIA,
 CC A CONDITION WHICH IS CHARACTERIZED BY ONE OR MORE OF THE
 CC FOLLOWING: LOSS OF PIGMENTATION, REDUCED EYE SIZE, FAILURE OF
 CC SECONDARY BONE RESORPTION, REDUCED NUMBERS OF MAST CELLS AND EARLY
 CC ONSET OF DEAFNESS, AND WHICH GIVES RISE TO A NUMBER OF DIFFERENT
 CC PHENOTYPES.
 CC -1- SIMILARITY: TO OTHER MEMBERS OF THE MYC FAMILY OF HELIX-LOOP-HELIX
 CC TRANSCRIPTION FACTORS.
 DR EMBL: L22958; AAB47773.1; -;
 DR EMBL: Z23056; CAA80600.1; -;
 DR EMBL: U19874; AAC52155.1; -;
 DR EMBL: U19875; AAC52156.1; -;
 DR HSSP: P22415; JANA4.
 DR MGD: MGI:104554; Mltf.
 DR PFAM: PF00010; HLH; 1.
 KW Nuclear protein; Transcription regulation; Alternative splicing;
 KW DNA-binding; Phosphorylation; Disease mutation;
 KW DOMAIN
 FT 201 217 ARG/LYS-RICH (BASIC).
 FT 267 288 LEUCINE-ZIPPER (POTENTIAL).
 FT 218 260 HELIX-LOOP-HELIX MOTIF.
 FT 1 65 MTSRIIRLOOLMROMOEOREROOOLKQAOQMOQRAVS
 FT VARSPLIC
 FT (IN ISOFORM 2).
 FT 67 202 MISSING (IN MICROPTHALMIA WHITE SPOT).
 FT 241 246 MISSING (IN MICROPTHALMIA SPOTTED).
 FT VARIANT 242 267 ACIPTESEARALAKEROKDNHMI -> V (IN
 FT MICROPTHALMIA EYELESS-WHITE).
 FT VARIANT 267 267 I -> N (IN MICROPTHALMIA WHITE).
 FT VARIANT 271 271 R -> K (IN MICROPTHALMIA OAK RIDGE).
 FT VARIANT 272 272 MISSING.
 FT VARIANT 277 277 D -> N (IN MICROPTHALMIA VITILIGO).
 FT VARIANT 318 474 MISSING (IN MICROPTHALMIA CLOUDY-EYED).
 FT SEQUENCE 474 AA; 52991 MW; 8E464430 CRC32;
 SO
 Query Match 2.3%; Score 191; DB 11; Length 474;
 Best Local Similarity 38.8%; Pred. No. 1,036-08;
 Matches 31; Conservative 23; Mismatches 21; Indels 5; Gaps 4;
 Db 252 RALAKERO-KKDNNTLERRRRNINDIKELGTLIPKSNPDMMKNGTKILKASVDYIR 310
 Oy 275 RVQPKVEKVSANHAIRRTSINDIKELNLV-GEQAKL--NRSAVLRKSIDKIR 331
 Db 311 KLOREOORAK-DLENPOKRL 329
 Oy 332 DLORONHDKAELORLOREL 351
 RESULT 10
 ID P91527 PRELIMINARY; PRT: 498 AA.
 AC P91527;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE DE SIMILARITY TO A HELIX-LOOP-HELIX MOTIF.
 GN W02C12.3.

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabdilita;
 OC Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE; 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., RULTON I.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., PULTON I.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAM J., KRISTEN J., LAISTER N., LATREILLE P.,
 RA LIGHNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTECK L., WILKINSON-SPROUT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA MURRAY J., WOHLDMAN P.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM H.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U80815; AAB37997.1; -;
 DR HSSP: P36956; AM9.
 DR PFAM: PF00010; HLH; 1.
 DR SEQUENCE 498 AA; 54391 MW; EAE8A806 CRC32;
 SO
 Query Match 2.3%; Score 188; DB 5; Length 498;
 Best Local Similarity 39.1%; Pred. No. 1,056-09;
 Matches 36; Conservative 25; Mismatches 26; Indels 5; Gaps 4;
 Db 238 KKDINMLERRRRNINDRIKELGOMLPKNTSEDKLKKGTILKASCYTRYLQKDEQCA 297
 Oy 284 KRSANHAIEKRRYRISINDIKEL--KNLV--VGEQAKLNKSAVLRKSIDKIRLORO-NHD 339
 Db 298 MKTQOQKSLSTAHKYNADRYKELEEMAROG 329
 Oy 340 LKAELORLQRELMARD-GSKVKDLQGLTRPG 370
 RESULT 11
 ID 075030 PRELIMINARY; PRT: 520 AA.
 AC 075030;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE A-TYPE MICROPTHALMIA ASSOCIATED TRANSCRIPTION FACTOR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RX MEDLINE; 98321192.
 RA AMAE S., FUSE N., YASUMOTO K., SATO S., YAJIMA I., YAMAMOTO H.,
 RA UDONO T., DURLU Y.K., TAKAI M., TAKAHASHI K., SHIBAHARA S.;
 RT "Identification of a novel isoform of microphthalmia-associated
 RT transcription factor that is enriched in retinal pigment epithelium."
 RL Biochem. Biophys. Res. Commun. 247:710-715(1998).
 DR EMBL: AB006909; BAA32288.1; -;
 DR HSSP: P22415; JANA4.
 DR PFAM: PF00010; HLH; 1.
 DR SEQUENCE 520 AA; 58162 MW; F5D32118 CRC32;
 SO
 Query Match 2.3%; Score 182; DB 4; Length 520;

```

Db      298 RALAKERO-KCDNNHLERRRRRINDRIKGLTLPKSNDPDMRWKKGITTLKASVDYIR 356
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      275 RVQRPVKVKKRSANNALERRYRTSINDIKNELKNLV-GEQAKL-NKSAVLKRKSIDIR 331

Db      357 KLOREOORAK-ELENROKRL 375
        ||||::|::|::|::|::|
QY      332 DLQRNHDLKAELQRLQREL 351

RESULT   12 PRELIMINARY; PRT: 793 AA.
AC       043019
DT       01-JUN-1998 (TREMBLrel.06, Created)
DT       01-JUN-1998 (TREMBLrel.06, Last sequence update)
DT       01-NOV-1999 (TREMBLrel.12, Last annotation update)
DE       HYPOTHETICAL 86.8 KD PROTEIN.
GN       SPBC354.05C.
OS       Schizosaccharomyces pombe (fission yeast).
OC       Eukaryota; Fungi; Ascomycota; Archiascomycetes;
CC       Schizosaccharomycetales; Schizosaccharomycetaceae;
NC       Schizosaccharomyces.
RN       [1]
RM       SEQUENCE FROM N.A.
RC       STRAIN=972H-.
RA       WOOD V., RAJNIDREAM M.A., BARRELL B.G., HILBERT H., DUESTERHOEFT A.;
RL       Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR       EMBL: AL022071: CAA17805.1: -.
DR       HSSP: P36956: 1AM9.
DR       PFM: PF00010: HLH: 1.
KW       Hypothetical protein.
SQ       SEQUENCE 793 AA; 86834 MW; 709C33AB CRC32;

Query Match          2.1%; Score 166; DB 3; Length 793;
Best Local Similarity 37.5%; Pred. No. 1,22e-06;
Matches 30; Conservative 27; Mismatches 18; Indels 5; Gaps 5;

Db      374 GPSDPSSTSVGKSYSDNSPISSATFAIQNGEPVGLSTOEONLSPLSKRSAHM 433
        ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      236 SPPVAPPTSGSRAS-KVRVAFLAPSPA-AMEVGKVPIN-RVQ-PKVKEV-KRSAHA 290

Db      434 IEKRYSLNDINKIAELRDAY 453
        ||::|::|::|::|::|::|
QY      291 IERRYRTSINDIKNELKNLV 310

RESULT   13 PRELIMINARY; PRT: 3247 AA.
AC       065553
DT       01-NOV-1996 (TREMBLrel.01, Created)
DT       01-NOV-1996 (TREMBLrel.01, Last sequence update)
DT       01-NOV-1999 (TREMBLrel.12, Last annotation update)
DE       UL36.
GN       UL36.
OS       Bovine herpesvirus 1.
OC       Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
CC       Alphaherpesvirinae; Varicellovirus.
RN       [1]
RM       SEQUENCE FROM N.A.
RC       STRAIN=COOPER;
RA       SCHWARTZ M., VLICK C., LOWERY D.E., BELLO L.J., MEYER G., MISRA V.,
RA       THURY E., PACES V.;
RL       Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN       [2]
RM       SEQUENCE OF 1-179 FROM N.A.
RC       STRAIN=COOPER;
RA       SCHWARTZ M., STYGER D., YOST B., LOWERY D.E., SIARD C.,
RA       LAOISSIERE S., MISRA V., VLICK C., PACES V.;
RA       Vel. Microbiol. 0:0-0(1996).
RN       [3]
RM       SEQUENCE FROM N.A.
```

[illegible]

Db 144 DMRNKGTLKASVDYIKWLQKEOORAR-ELEHROKRL 180
 QY 316 KL--NKSAVLRKSIDKIPDLORONHDLKAELOREL 351

RESULT 15
 ID 014948 PRELIMINARY; PRT; 347 AA.
 AC 014948;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
 DE TREC ISOFORM (OR TREC).
 GN TREC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MONOCYTIC LEUKEMIA;
 RX MEDLINE: 97398136.
 RA YASUMOTO K., SHIBAHARA S.;
 RT "Molecular cloning of cDNA encoding a human TREC isoform, a newly
 identified transcriptional regulator";
 RL Biochim. Biophys. Acta 1353:23-31(1997).
 DR EMBL: D43945; BAA21908.1; -
 DR HSSP: P22415; IAN4.
 DR PFAM: PF00010; HLH: 1.
 SO SEQUENCE 347 AA; 38787 MW; C2C3DFF6 CRC32;

Query Match 2.0%; Score 159; DB 4; Length 347;
 Best local Similarity 31.6%; Pred. No. 1.07e-05;
 Matches 30; Conservative 30; Mismatches 30; Indels 5; Gaps 4;

Db 117 PSSLPMKREITETTRALAKERO-KDQNHLLIERRRINRYRIKELGTLIPKSNDDPMR 175
 QY 260 PSPAPMEVQGVPIINRVQPKYKVSANNAIERRRYRISINDKINELKLVY-GEQAKL- 317
 Db 176 MNKGTLKASVEYIKWLQKEOORAR-ELEHROKRL 209
 QY 318 -NKSAVLRKSIDKIPDLORONHDLKAELOREL 351

Search completed: Sun Sep 3 12:50:47 2000
 Job time : 163 secs.

THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

WSEDF
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:46:39 2000; Maspar time 34.35 Seconds

Tabular output not generated. 1004.075 Million cell updates/sec

Title: >US-09-332-522B-8
Description: (1-1113) from US09332522B.pep
Perfect score: 8012
Sequence: 1 MDITLNLNLDAPLDESMDFL.....LKECYQLMKSLGSGSVKA 1113

Scoring table:
PAM 150
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-Prot38
1:swissprot

Statistics: Mean 57.220; Variance 124.660; scale 0.459

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1494	18.6	1147	1	SREL_HUMAN STEROL REGULATORY ELEM	1.65e-247
2	1418	17.7	1133	1	SREL_CRIGR STEROL REGULATORY ELEM	6.26e-233
3	1278	16.6	1141	1	SRE2_HUMAN STEROL REGULATORY ELEM	1.05e-215
4	1278	16.0	1139	1	SRE2_CRIGR STEROL REGULATORY ELEM	3.68e-206
5	1234	15.4	1023	1	SREL_RAT STEROL REGULATORY ELEM	8.99e-198
6	192	2.4	446	1	TEF3_MOUSE TRANSCRIPTION FACTOR E	2.41e-10
7	190	2.4	743	1	TEF3_HUMAN TRANSCRIPTION FACTOR E	4.45e-10
8	162	2.0	514	1	TEF3_HUMAN TEF3 PROTEIN (FRAGMENT	3.39e-06
9	149	1.9	160	1	MAX_HUMAN MAX PROTEIN.	1.67e-04
10	149	1.9	160	1	MAX_HUMAN MAX PROTEIN.	1.67e-04
11	149	1.9	160	1	MAX_HUMAN MAX PROTEIN.	1.67e-04
12	149	1.9	201	1	YAMC_SCHPO HYPOTHETICAL 23.0 KDA	1.67e-04
13	151	1.9	474	1	VTP3_TTVIV VIRAL PROTEIN TPX	9.27e-05
14	150	1.9	582	1	MNT_HUMAN MAX BINDING PROTEIN MN	1.24e-04
15	150	1.9	817	1	VRP1_YEAST VERPROLIN.	1.24e-04
16	149	1.9	2414	1	P300_HUMAN ELA-ASSOCIATED PROTEIN	1.67e-04
17	142	1.8	160	1	MAX_MOUSE MAX PROTEIN (MYN PROTE	1.26e-03
18	148	1.8	163	1	MAX_XENLA MAX PROTEIN (MAX)	2.24e-04
19	142	1.8	422	1	MYC_AYIMC MYC TRANSFORMING PROTE	1.26e-03
20	148	1.8	591	1	MNT_MOUSE MAX BINDING PROTEIN MN	2.24e-04
21	147	1.8	620	1	EXTN_TOBAC EXTENSIN PRECURSOR (CE	7.13e-04
22	144	1.8	865	1	CPN_DROME CALPHOTIN.	7.13e-04
23	141	1.8	1586	1	SN22_HUMAN POSSIBLE GLOBAL TRANSC	1.68e-03

24	133	1.7	310	1	USF1_HUMAN	UPSTREAM STIMULATORY F	1.57e-02
25	133	1.7	310	1	USF1_RABIT	UPSTREAM STIMULATORY F	1.57e-02
26	133	1.7	310	1	USF1_MOUSE	UPSTREAM STIMULATORY F	1.57e-02
27	136	1.7	364	1	HK61_MESAU	HOMEOBOX PROTEIN NKX-6	6.85e-03
28	136	1.7	364	1	HK61_RAT	HOMEOBOX PROTEIN NKX-6	6.85e-03
29	137	1.7	367	1	HK61_HUMAN	HOMEOBOX PROTEIN NKX-6	6.85e-03
30	133	1.7	378	1	HAIR_DROVI	HAIR PROTEIN	5.18e-02
31	134	1.7	422	1	MYC_AYIMD	MYC TRANSFORMING PROTE	1.19e-02
32	133	1.7	434	1	NO75_LUPLU	EARLY NODULIN 75 PROTE	1.57e-02
33	136	1.7	454	1	MYC2_MARMO	N-MYC 2 PROTO-ONCOGENE	6.85e-03
34	133	1.7	454	1	MYC2_SPEBE	N-MYC 2 PROTO-ONCOGENE	1.57e-02
35	135	1.7	464	1	SP62_HUMAN	SPICEOSOME ASSOCIATED	9.03e-03
36	135	1.7	583	1	BMAL_HUMAN	BMAL1 PROTEIN (BRAIN A	9.03e-03
37	133	1.7	700	1	B1B_DROME	NEUROGENIC PROTEIN BIG	1.57e-02
38	140	1.7	822	1	EP58_HUMAN	EPIDERMAL GROWTH FACTO	2.23e-03
39	135	1.7	822	1	Y539_HUMAN	HYPOTHETICAL PROTEIN K	9.03e-03
40	135	1.7	939	1	S720_YEAST	SERINE/THREONINE-PROTE	9.03e-03
41	137	1.7	2038	1	FSH_DROME	FEMALE STERILE HOMEOTI	5.18e-03
42	131	1.6	331	1	MAZ_MESAU	MYC-ASSOCIATED ZINC FI	2.70e-02
43	132	1.6	423	1	MYC_AYIME	MYC TRANSFORMING PROTE	2.06e-02
44	132	1.6	429	1	MYC5_RAT	S-MYC PROTO-ONCOGENE P	2.06e-02
45	132	1.6	721	1	YK82_MYCTU	HYPOTHETICAL 73.6 KDA	2.06e-02

ALIGNMENTS

RESULT 1
ID SREL_HUMAN STANDARD: PRT: 1147 AA.
AC P36936;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE STEROL REGULATORY ELEMENT BINDING PROTEIN-1 (SREBP-1) (STEROL
DE REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 1).
GN SREBP1 OR SREBP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM H.A., AND PARTIAL SEQUENCE.
RX MEDLINE: 94006541.
RA Yokoyama C., Wang X., Briggs M.R., Admon A., Wu J., Hua X.,
RA Goldstein J.L., Brown M.S.;
RT "SREBP-1, a basic-helix-loop-helix-leucine zipper protein that
RT controls transcription of the low density lipoprotein receptor
RT gene.";
RT Cell 75:187-197(1993).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 319-394.
RX MEDLINE: 98298442.
RA Parraga A., Belisolel L., Ferre-D'Amare A.R., Burley S.K.;
RT "Co-crystal structure of sterol regulatory element binding protein 1a
RT at 2.3-A resolution.";
RL Structure 6:661-672(1998).
CC -1- FUNCTION: "TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE STEROL
CC REGULATORY ELEMENT 1 (SRE-1) (5'-ATCAGCCGAC-3') FOUND IN THE
CC PLANKING REGION OF THE LDLR GENE AS WELL AS OTHER GENES.
CC ADD1/SREBP1 HETERODIMER HAS DUAL SPECIFICITY, BINDING TO
CC BOTH AN E-BOX MOTIF (ATCAGCGA) AND TO SRE-1.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC B1H PROTEIN.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANES OF THE NUCLEAR
CC ENVELOPE AND ENDOLASMIC RETICULUM. RELEASED INTO THE NUCLEUS UPON
CC PROTEOLYTIC CLEAVAGE.
CC -1- ALTERNATIVE PRODUCTS: SEVERAL ISOFORMS DIFFERING IN THE N- OR C-
CC TERMINI ARE PRODUCED BY ALTERNATIVE SPLICING. THE SEQUENCE SHOWN
CC IS THAT OF SREBP-1A.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN A WIDE VARIETY OF TISSUES. MOST
CC ABUNDANT IN LIVER AND ADRENAL GLAND. IN FETAL TISSUES LONG AND
CC LIVER SHOWS HIGHEST EXPRESSION.
CC -1- PTM: UNDER STEROL-DEPLETED CONDITIONS, SREBPs ARE PROTEOLYTICALLY
CC CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS
CC AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE

CC	PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,				
CC	INDEPENDENT OF STEROL LEVELS.				
CC	-1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF				
CC	TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.				
CC	-----				
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	at the European Bioinformatics Institute. There are no restrictions on lists				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/sib-ch).				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: U00968: CAB34682.1; -.				
DR	PDB: 1AM9; 01-JUL-98.				
DR	TRANSFAC: T01556; -.				
DR	MIM: 184756; -.				
DR	PFAM: PF00010; HLH: 1.				
DR	PROSITE: PS00038; HELIX-LOOP-HELIX: 1.				
KW	Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;				
KW	Endoplasmic reticulum; Activator; Apoptosis; 3p-structure.				
FT	DOMAIN 1 60				
FT	DOMAIN 61 178				
FT	BINDING 92 416				
FT	-----				
FT	DNA_BIND 323 336				
FT	DOMAIN 337 374				
FT	DOMAIN 373 394				
FT	DOMAIN 427 462				
FT	SIZE 460 461				
FT	-----				
FT	TRANSMEM 488 508				
FT	TRANSMEM 548 568				
SO	SEQUENCE 1147 AA; 121644 MW; 58BD93A2371F3E9 CRC64;				
	POTENTIAL.				

	BASIC DOMAIN.				
	HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).				
	LEUCINE-ZIPPER (POTENTIAL).				
	GLY/PRO/SER-RICH.				
	CLEAVAGE (BY APOPAIN AND CASPASE-7) (BY				
	SIMILARITY).				
	POTENTIAL.				

Query Match	18.6%;	Score 1494;	DB 1;	Length 1147;
Best Local Similarity	33.6%;	Pred. No. 1.65e-247;		
Matches	338;	Conservative	227;	Mismatches 385;
			Indels	55;
			Gaps	51;

```

Db 169 PGSTGSPCNTQOPLRGLRGLASPRGVPVLSLTQ-V-OSVAVQOULLTJTAAPTAAPV 226
Oy 126 PTGLKKAQOPATIHMDAQRMPPRTAYAPPSLSSFFYOSMSPPIS-PYESAONONYNM 184
Db 227 TTVTSQIOQVSVLLQPHFIKADSLSLTAMTKDGAFTVAAGLSPLVSGTQTQGLPTLV 286
Oy 185 QPVATAPASAPLPQOSY-PQPFITYNSKA-GWTSDEA-MYLLQGTVASPPSPVPV 241
Db 287 GGTILATVPLVYDAEKL-PT-NRLAAGSKAPAS-AOSRG-E-KTANMAIEKTRSSIND 341
Oy 242 PPTSGSRASRVAPALPASPAAEVQGVKYNVQPVKVRKSANHAIEKRRYSIND 301
Db 342 KIIEKIDLVYCTEAKLINKSAVLRAIDYIFRLOHSONKLOENSL-RTVHWKSKS-LKD 399
Oy 302 KINEKKNLVYVEQAKLNKSAVLKRSINDIKRLOQONHDLKELORLREIMARQSKVKD 361
Db 400 LVS-ACSGSGNTDVLMEGVKT-EVEDTLPPSDAGSPFOSSPLSLGSRSGSGSGSDS 457
Oy 362 LLOGTGRGRASKRRRESSQFTTDDAGTLPPRSDSDPSLS-PMH-SDISLPPVGGST 419
Db 458 EPDSEVFEDSAKPEQORSLSRGLMDBSRLACTLYFLCISCPNLASLIGARLPSPSD 517
Oy 420 ASCSGSSSSNEEPLVYVPS-EMRGMACTHSRGLGCMFMALIVANPFTFL-OROHYSDND 477
Db 518 TTSVYHSGRNVLYGESRDGPMQOMLLPVMVLLNGLVLYSVLLFYVGEPTPRHSG 577
Oy 478 DLGDMSGO-RILSLDV-EGEGFAVWOOSWIMLLNFTLMIGCYKLVLLIGDPOLDQTD 535
Db 578 PAVYFMRRKQADDLIARGDPAQAAOQLTML-LRALGRPLPTSHDLACSL-L-WNLIRH 634
Oy 536 -A-YC-QHRORADYFVSGGGSQSAVAG-YLNCIMHFGISPAUSLE--CYLQTTWQRLR 589
Db 635 LLQRLVGRWMLAGRAGCLOQDCALFVDASASARPAALVYHKLHOLHTMGKHTGHTATN 694

```

QY	590	LFEHLMVGRVLSRSRSGLEFSPNNAASRKOALASARELALFNRLNQLQJLNGNSRQDMNGM	649
Db	695	LALSALNLEACGADVASVTAETAEIYYAALARKTSLPALH-FLTRFLLSSARQAC-LAQ	752
QY	650	MAFLPASMVAEVAHNLITPRETICIHMTALRRKRSPAPWIDOFARYYMSARQECGTR	709
Db	753	SGSVPRAMQWLCHPVGRHFRFVODGMSV-LSTP--WESLYSLAGNVDPDLAQVTLFREHL	809
QY	710	ATEGTQELRRALFAFYGRICATHVFETLSDSGEDDGGFTLRNRNCPDPAHYIKQYREHL	769
Db	810	LERALNCTVTPNPSPGADGKEFSDALGYLIOLLNSCSDAAGAPAFSPSSISM-ATTG	868
QY	770	LFSISQVLVAGHKSGGLPTSSVSGAEQLQQQHSSTIVSNLYKTSLLADTLMADE	829
Db	869	VDPYAKWASLIVAVIHTMRDEBAERLCPLVEHLPRYLDE-SER-PLPRAALHS-FKA	925
QY	830	RDTWVWMAADVLETAHVMTLGEDPTLAEOLYGRIKOMPTLOCGENDHLPKA-LHAYLRA	888
Db	926	ARALL-GCAAA-BSGPASLT-I-CERASGYLODLSATTPASSID-KAV-OLFLQDILLY	979
QY	889	KMILLKNGNALDKSLKQLVNITCDSSVELOECLTVNRITDAKGIKLLFOLLTCDWLE	948
Db	980	VRTISFROQPPAPAPAAOGASSRPGASALELRCFORDLSLRLAQSFPRAMRVFLHE	1039
QY	949	TRTALM-ELEHNM-M-EDDGFYQVP-GEVLE-K-FQTDLSNRINVENIPMAQSRIVYE	1002
Db	1040	ATATLMAGASPTFHQLDLRSLLRRRAGPGKG-GAVAELEPR-PTREHAEALLASCY	1096
QY	1003	AVCLIMAGASPCPTQOULDRSLRSFNAHSSIFCGKDRQONFVGGERERASAMYACKY	1062
Db	1097	LPRGFLSAPGQRYGMLEAARTLEKIGDRRLHDDOQMLRGGG	1141
QY	1063	LPALLSSPGERAGMLAEAKTLEKVGDRKRKICEYQJLMSLGG	1107
RESULT	2		
ID	SREL CRIGR	STANDARD:	PRT: 1133 AA.
AC	Q60416,		
DT	15-JUL-1998 (rel. 36, Created)		
DT	15-JUL-1998 (rel. 36, Last sequence update)		
DT	15-FEB-2000 (rel. 39, Last annotation update)		
DE	STEROL REGULATORY ELEMENT BINDING PROTEIN-1 (SREBP-1) (STEROL		
DE	REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 1).		
GN	SREBP1 OR SREBP1.		
OS	Citricellus griseus (Chinese hamster).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;		
OC	Citricellus.		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 94274723.		
RA	Sato R., Yang J., Wang X., Evans M.J., Ho Y.K., Goldstein J.L.,		
RA	Brown M.S.;		
RT	Assignment of the membrane attachment, DNA binding, and		
RT	transcriptional activation domains of sterol regulatory		
RT	element-binding protein-1 (SREBP-1). ¹		
RL	J. Biol. Chem. 269:17267-17273(1994).		
CC	-1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE STEROL		
CC	REGULATORY ELEMENT 1 (SRE-1) (5'-ATCACCAC-3') FOUND IN THE		
CC	FLANKING REGION OF THE LDLR GENE AS WELL AS OTHER GENES.		
CC	ADD1/SREBP1 HETERODIMER HAS DUAL SEQUENCE SPECIFICITY, BINDING TO		
CC	BOTH AN E-BOX MOTIF (ATCAGCTGA) AND TO SRE-1.		
CC	-1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER		
CC	B1H PROTEIN.		
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANES OF THE NUCLEAR		
CC	ENVELOPE AND ENDOPASMIC RETICULUM. RELEASED INTO THE NUCLEUS UPON		
CC	PROTEOLYTIC CLEAVAGE (BY SIMILARITY).		
CC	-1- PTM: UNDER STEROID-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY		
CC	CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS		
CC	AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE		
CC	PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,		
CC	INDEPENDENT OF STEROL LEVELS.		
CC	-1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF		

RESULT	2	STANDARD:	PRT:	1133 AA.
ID	SREL CRIGR			
AC	060416:			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	STEROL REGULATORY ELEMENT BINDING PROTEIN-1 (SREBP-1) (STEROL			
DE	REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 1).			
GN	SREBP1 OR SREBP1.			
OS	Citellus griseus (Chinese hamster).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;			
OC	Citellus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 94274723.			
RA	Sato R., Yang J., Wang X., Evans M.J., Ho Y.K., Goldstein J.L.,			
RA	Brown M.S.;			
RT	Assignment of the membrane attachment, DNA binding, and			
RT	transcriptional activation domains of sterol regulatory			
RT	element-binding protein-1 (SREBP-1).";			
RL	J. Biol. Chem. 269:11267-11273(1994).			
CC	-1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE STEROL			
CC	REGULATORY ELEMENT 1 (SRE-1) (5'-ATCACCCAC-3') FOUND IN THE			
CC	ADDI PLANNING REGION OF THE LDLR GENE AS WELL AS OTHER GENES.			
CC	ADD1/SREBP1 HETERODIMER HAS DUAL SEQUENCE SPECIFICITY, BINDING TO			
CC	BOTH AN E-BOX MOTIF (ATCACGTCG) AND TO SRE-1.			
CC	-1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER			
CC	BHLH PROTEIN.			
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANS OF THE NUCLEAR			
CC	ENVELOPE AND ENDOPLASMIC RETICULUM, RELEASED INTO THE NUCLEUS UPON			
CC	PROTEOLYTIC CLEAVAGE (BY SIMILARITY).			
CC	-1- PTM: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY			
CC	CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS			
CC	AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE			
CC	PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,			
CC	INDEPENDENT OF STEROL LEVELS.			
CC	-1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF			


```

RU J. Biol. Chem. 270:12152-12161(1995)).
CC -FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE STEROL
CC REGULATORY ELEMENT 1 (SRE-1) (5'-ATCACCCAC-3') FOUND IN THE
CC FLANKING REGION OF THE LDLR AND HMG-COA SYNTHASE GENES (BY
CC SIMILARITY).
CC
CC -SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN..
CC
CC -SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANES OF THE NUCLEAR
CC ENVELOPE AND ENDOPLASMIC RETICULUM. RELEASED INTO THE NUCLEUS
CC UPON PROTEOLYTIC CLEAVAGE.
CC
CC -PTM: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY
CC CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS
CC AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE
CC PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,
CC INDEPENDENT OF STEROL LEVELS.
CC
CC -DISEASE: STEROL-RESISTANT DEFECTIVE (SRD) PHENOTYPES EXPRESS
CC TRUNCATED FORMS OF SREBP-2 PROTEIN, OFTEN SREBP-2 IS FUSED TO OTHER
CC PROTEINS, AS IS THE CASE IN SRD-1, WHERE SREBP-2 IS FUSED TO AN
CC OUT-OF-FRAME KU P70 PROTEIN OR, IN SRD-2 WHERE THE FUSION PROTEIN
CC IS A LIM DOMAIN-CONTAINING PROTEIN. SRD PHENOTYPES ARE RESISTANT
CC TO STEROL BIOSYNTHESIS REPRESSION BY STEROIDS.
CC
CC -SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUPERFAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL: U12330: AAA74141.1: -
CC EMBL: U12329: AAA74140.1: ALT_TERM.
CC EMBL: U22819: AAA85719.1: ALT_TERM.
CC EMBL: U22818: AAA85718.1: ALT_TERM.
CC HSSP: P36956: 1A99.
CC PFAM: PF00010: HLH: 1.
CC PROSITE: PS00038: HELIX-LOOP-HELIX: FALSE NEG.
CC Transcription regulation: DNA-binding; Nuclear protein; Transmembrane;
CC Endoplasmic reticulum; Polymorphism; Chromosomal translocation.
CC
CC DDOMAIN 1 50
CC DDOMAIN 51 120
CC DDOMAIN 121 244
CC BINDING 95 421
CC
CC TO STEROL REGULATORY ELEMENT-1 (BY
CC SIMILARITY).
CC
CC BASIC DOMAIN.
CC HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
CC LEUCINE-ZIPPER (POTENTIAL).
CC BREAKPOINT FOR TRANSLLOCATION TO FORM
CC SREBP-2 FUSION PROTEINS IN SRD
CC PHENOTYPES.
CC
CC CLEAVAGE (BY APOPAIN AND CASPASE-7) (BY
CC SIMILARITY).
CC
CC POTENTIAL.
CC POLY-ALA.
CC POLY-SER.
CC S-> N (IN 50% OF THE MOLECULES).
CC
CC SEQUENCE 1139 AA: 123655 MW: E81C2778BEF02653 CRC64;

```

ID	RESULT	5	STANDARD;	PRT;	1023	AA.
Db	224	ATVOTVAPRVQVYVLPVQRIKITSLSVLTITKTGSPVMAAVONPALTALTPLQTA	283			
Qy	187	--VAATPAPASAPLPQOSYPOPFITYN-SKAGMTSDSEMYLT-LQ-PTVAS--PTSPSPA	240			
Db	284	LQVPLVGSNQTILTPVMVGGKVPYIKQVPGCVQVQLEPPKCGERTNHNIEKRYSS	343			
Qy	241	-PPPTSGTASAKRVAPLAPSPAMEV-OGKVPINVOQKVEYKRSAINALERTYRFS	298			
Db	344	INDKIIEIKLDVMQTDAMKMRSGYLKRAIDYIYVLOOVNFKLEQKEMVU-K-LANQKRL	401			
Qy	299	INDKINELKNIYVEQAKLKNSAHLKRSIDKINDLQRONHDLKAEIQRORLQRELMADGSK	358			
Db	402	LKGI-DIGSLVSDVDLKI-E-D-FNONVLLMSPASDSG-SQAGF-SPYSIDSEP--GS	453			
Qy	359	VKDLQITQHPGRASKRRRESSQFTTDALTPRPSDESDPSLSPMHSIDSLPPSYGGS	418			
Db	454	PLL---DDAKYKDEPDSPPAL-GYDRSKITLCVTLFLGSLSTPLSLQMGCHDQD	509			
Qy	419	TASCGSSSSSSNEPDLVPPSMRMATHSRLGCMFALANPKPTFLQRHGYSDND	478			
Db	510	HPYSGSGRSTYLSLEGSGGWFDMMPILMLIANGVYLSVEFKLVHGEPTVLRPTSS	569			
Qy	479	LGDMSGQRITISTYVEEGEGRVAWQSSWI-WLNFITLMGCLVKLVLYGDPQIDAOT--D	535			
Db	570	VTFMRHKQADLDIARGFMAAANLQTCLSVLGRALPTSRDLASCL--SWNVIRYSIQ	627			
Qy	536	-AYCGHQRADFYFSQGSQSSAVAGYINCLHMGSLSPARLE--CYLQDTWQFLFLPH	592			
Db	628	KLRILVRILKKVFORMRATPATAGFEDEKSSARDAALAHRLHOLHTGKLPAGSTGS	687			
Qy	593	--RL-W-LGRVLSRRSGGLEFSMAASKQALASARELALFNRLNLOLQMGNSRQDM-N	646			
Db	688	DVHAALCAVNLAECAEKIPRSTLIEHLAAAGLKTGCGGKIG-PLASYFLRAQSLG	746			
Qy	647	GIMMALTRSMAEVAHNLIPRETICHHVTFILPMKRSAPKMLQGFARYMSRAQEGC	706			
Db	747	PEHSTVP-DSLRWICHPLGOKFEWER--SWSIK-SAKESLYCAQRNPADPIAQVHAFC	802			
Qy	707	RTRAEDTQOELRMFAFTAYGRYCATIHFTYDLSGSDGQDQFTRLNPPCPAHVIKOYR	766			
Db	803	KNLEPRAVESLYKROSKKSGDDECEFSALLETIKLHSHVSDVSGVTPPFSSSVL	862			
Qy	767	EHLHFKSIQCLVGA-GHKSGGLPTSSVSGAEOLQOQOHSGITVSNV-LKYSILDLTIM	824			
Db	863	KSALGPDIKRMWTSAYTMATISMLQGDAAVNRSHFEVEEIPKALE-VNES-PLYKAVY	920			
Qy	825	ADEDEBDTNYVWADVLETAIVHMLGEDTLAEOLYGRIKOMPOLQOCGNDHLPALHA	884			
Db	921	ACRAMHASLSGCKGQONSF---CH-CERASGHLWMSLVNSGATSDPSLNHVQLTCD	975			
Qy	885	VLRKMKILLKNGNALDLKSLKOLVNLICDSSVELQGLVNNITPAKGIKLLFOLLTCD	944			
Db	976	LLLSLRPALWOKGASQSLGETYHASGTETIAGFORDJGSLRPLASHFRPARYKVFLEP	103			
Qy	945	WLLETFATLWELHNMNEDG-FYOVPGVELFEFQDNLNINVENIPAOGRIVLYEA	100			
Db	1036	TIVRLACANITRTTHQLLEHSLRRITSQNTNH-GEIDTWP-----GQERATALLACRHL	108			
Qy	1004	VCRLMAAGSCPQOQLDLDRSLRSNAHSITFCGSKORPOONFYVGERRRSANYVACKYL	106			
Db	1090	PLSFSSPGGRAVLLAAARLTLEKVGDRSCNCOOMIYVLAGGTA	1135			
Qy	1064	PRALLSPGRAGMLAANAATLEKVGDRKRLKCYQLMKSLGNGIG	1109			
RESULT	5	STANDARD;	PRT;	1023	AA.	
ID	SREL_RAT					
AC	P56720;					
DT	15-FEB-2000 (Rel. 39, Last sequence update)					
DT	15-FEB-2000 (Rel. 39, Last annotation update)					
DE	STEROL REGULATORY ELEMENT BINDING PROTEIN-1 (SREBP-1) (STEROL					
DE	REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 1) (ADIPOCYTE					

DE DETERMINATION-AND DIFFERENTIATION-DEPENDENT FACTOR 1) (ADD1)
 DE (FRAGMENT).
 GN SREBP1 OR SREBP1.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ADIPOCYTE.
 RX MEDLINE: 93330269.
 RA Tortonoz P., Kim J.B., Graves R.A., Spiegelman B.M.;
 RT "ADP1: a novel helix-loop-helix transcription factor associated with
 RT adipocyte determination and differentiation.";
 RL Mol. Cell. Biol. 13:4753-4759(1993).
 RN [2]
 RP DNA-BINDING.
 RX MEDLINE: 95257939.
 RA Kim J.B., Spotts G.D., Halvorsen Y.D., Shih H.M., Ellenberger T.,
 RA Towle H.C., Spiegelman B.M.;
 RT "Dual DNA binding specificity of ADP1/SREBP1 controlled by a single
 RT amino acid in the basic helix-loop-helix domain.";
 RL Mol. Cell. Biol. 15:2582-2588(1995).
 CC -1- FUNCTION: REGULATED DURING BOTH ADIPOCYTE DETERMINATION AND
 CC DIFFERENTIATION. ADP1/SREBP1 HETERODIMER HAS DUAL SEQUENCE
 CC SPECIFICITY, BINDING TO BOTH AN E-BOX MOTIF (ATCAGTGA) AND TO
 CC SRE-1 (5'-ATCAGCCAC-3').
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
 CC BHLH PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS; SREBP-1A AND SREBP-1C/ADP1
 CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BROWN ADIPOSE
 CC TISSUE.
 CC -1- PTM: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY
 CC CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS
 CC AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE
 CC PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,
 CC INDEPENDENT OF STEROL LEVELS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION, MANY PROBABLE
 CC FRAMESHIFTS WERE CORRECTED FROM POSITION 878 ONWARD.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L16995; NOT ANNOTATED. CDS.
 DR PROSITE: PS00038; HELIX LOOP HELIX: 1.
 KW Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;
 KW Endoplasmic reticulum; Apoptosis; Alternative splicing.
 FT BINDING 67 385
 FT DOMAIN 125 128
 FT DNA_BIND 293 306
 FT DOMAIN 307 344
 FT DOMAIN 343 364
 FT SITE 427 428
 FT TRANSMEM 454 474
 FT TRANSMEM 513 533
 FT NON_TER 1023 1023
 SO SEQUENCE 1023 AA; 108571 MW; 2593DF46B2A11C8F CRC64;
 Query Match 15.4%; Score 1234; DB 1; Length 1023;
 Best Local Similarity 32.1%; Pred. No. 8,99e-198;
 Matches 290; Conservative 218; Mismatches 340; Indels 56; Gaps 50;

QY 147 MPPNTAVYVPSLGSSFFVYQSMSPPTSPVESANONVNMQVAAAPAP-ASAPLPQ-QSYP 204
 DB VYLOPFIKADSLTLTAVKTPDGTATKTAGINTLAPATAVOAGPLOTVSGTITATVPL 266
 QY 205 ---QP-FITYSKA-GMTSDAMYLLOPTYASPPSPVAPPRTSTGSRASKRYVA-PL 258
 DB 267 VVDIDKLPIDHRLAAGKALGSAQSGEKRATANAIEKRRSSINDKIVELDLVVGTEAK 326
 QY 259 APSPAVEVQGVKPIINR-V-QPKVEYKRSANHAITERRTSINDKINELKNLVGEOAK 316
 DB 327 LNKSAVLKALDITRFLOHSNOKLKOENLTL-RS-AKXSLNDLVY-ACSGGGTIVS 382
 QY 317 LNKSAVLKRSIDKTRDLORHNDLKAELQRIQRELMADGSKYVDLQLGRRPGASAKR 376
 DB 383 MEGMKPEYVET-LTPPPDAGSPSSPSPLS-LGSRGSSGSGSDSPDPAPFEDNOVKQR 440
 QY 377 RESSQFTTIDAGLTTPRDESDPSLSPHSDISLPPSYGGSTASCGSSSSNE-EPVY 435
 DB 441 LPSHSGMLDPSRLALCVLFELCTCNPLASLFGWGLTTPDASGVRHRSRSMLEASR 500
 QY 436 VPSSRMGMATHSRLGCLGMFMAILAVNPFKTFLORGHDSNDLG-DMSGGRILSYDV- 493
 DB 501 DGSNMTQWLLPPLVWLANGLLVLAALLFYGEVYTRPHSGPAVHFWRKROADLALR 560
 QY 494 EGEGRVAVQSSWIMLNTFTLMLGCLVRLVYGDPOLDQRPD-A--YCOHGRADPFYSQ 550
 DB 561 GDPFAQAQWMLA-LQALGRPLTPSNLDACSLD--NVLVHLLQRLVWGRMLAGQAGL 617
 QY 551 GOSSQAYAG-YLNCILHMFGLSPASRLD-CYLQTTWFLFLFRLMLGKYSRSRSG 607
 DB 618 QRDYLRKRDARASADAAYVYHKLQHLAMGKYTGHLVASNALALNLAECAGDAIS- 676
 QY 608 FSNASRQALASARELLALFRLNLQDTGNGSRGDMGIMMALFASMAEVAHNLTP 667
 DB 677 MATLAEIYVAAALRKYSILPRALH-FLTRFLSSARQAC-LAOSGAVLAQWMLCHPVG 734
 QY 668 RRTIC-HVMYALNRKRRAKAPWLOOFFARYYSRROCGRTARAEQTOELRMATFAGY 726
 DB 725 RF-FVDG-DMAV-HGAPESISYVAGNPVDPLAQVTRFCEHLERALNCTAOPSPGAD 791
 QY 727 RLCATHVTYDLSGDEDDGFFTRLRNCDDRAHYIKRYREHLFKSTQCLVAGAKHSG 786
 DB 792 -GHREFSALGYQLLNSCSDAVGAPA-CSFSVSSM-ATTGTDPVAKWMASTIAVYH 848
 QY 787 LPTSSVSEAEQLOQOQSGITVSNVLTYSILKXTLWADEDERTNVWMAADVLETAVH 846
 DB 849 WLRBEAEARERYLVEHITPOVLOE-TER-PLPRALYS-EKAAALLDHR-K-VESSPA 903
 QY 847 WLGEDTLAEOLYGRIKOMPTOLOQCGENDHLPKA-LHAVLRKAMWILKNGNALDKSLK 905
 DB 904 SLA-I-CERKASGYLRDSTASTASSID-KAM-QLLCDLLLVARTSLXROOASASAGA 959
 QY 906 QLVNLCDESSVELOECITVNRITDAKIKILLFOLLTCDWLEFETAL-WELEHNMMD 964
 DB 960 HGTSGPOASALELRFQHDLSLRLAOSFRPAMRVFLHEATARLMAAGSPARTHLL 1019
 QY 965 -GFYGVV-GEVLE-K-FQDINSLRNIIVENIPNARSRIYEAQVRLMAGASPCPTOLL 1020
 DB 1020 DRGI 1023
 QY 1021 DRSL 1024
 RESULT 6
 ID TFE3 MOUSE STANDARD: PRT: 446 AA.
 AC Q64092;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE TRANSCRIPTION FACTOR E3 (FRAGMENT).
 GN TFE3 OR TCFE3.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AC P37370; Q06133;
DI 01-OCT-1994 (Rel. 30, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 01-NOV-1997 (Rel. 35, Last annotation update)
DE VERPROLIN.
GN VRP1 OR MDP2 OR ENDS OR YLR337W OR L8300.13.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A364;
RX MEDLINE; 95058201.
RA Donnelly S.F.H., Pocklington M.J., Pallota D., Orr E.;
RT "A proline-rich protein, verprolin, involved in cytoskeletal
RT organization and cellular growth in the yeast Saccharomyces
RT cerevisiae";
RL Mol. Microbiol. 10:585-596(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifken L., Riles L., Taich A., Trevasakis E., Vignati D.,
RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: INVOLVED IN CYTOSKELETAL ORGANIZATION AND CELLULAR
CC GROWTH. MAY EXERT ITS EFFECTS ON THE CYTOSKELETON DIRECTLY, OR
CC INDIRECTLY VIA PROLINE-BINDING PROTEINS (E.G. PROFILIN) OR
CC PROTEINS POSSESSING SH3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL; Z26645; CAA81388.1; -
DR EMBL; 019028; AAB67263.1; -
DR PIR; S39626; S39626.
DR SGD; L0002481; VRP1.
KW Cytoskeleton; Repeat.
FT DOMAIN 5 14
FT DOMAIN 239 245
FT DOMAIN 349 357
FT DOMAIN 396 406
FT DOMAIN 424 431
FT DOMAIN 462 468
FT DOMAIN 704 708
FT CONFLICT 308 308
FT CONFLICT 350 350
FT CONFLICT 689 689
FT CONFLICT 710 817
FT
FT
FT
FT
SQ SEQUENCE 817 AA: 82593 MW: 24C7522D5B1CA1C8 CRC64: 1.
Query Match 1.98; Score 150; DB 1; Length 817;
Best Local Similarity 25.34; Pred. No. 1.24e-04;
Matches 46; Conservative 48; Mismatches 81; Indels 7; Gaps 7;
Db 293 SSTRKIOTENHKSPSPPLPSAP-P-IPITSHA-PPLPPTAPPSPSLPNVTSAPKKAISAP 349
QY 89 SPVHIKEELHQOQOQSPFLVYKRPDLIATSYNCPQOQPTGLAKAQPATITIHMDAQRMP 148
Db 350 APPPPLPAMSSASTNSVKATPVPTLAPPLDPTTSVPPNKASSMPAPPPPPP-PGA 408

QY 149 -ENTAVYPPSLSSFEYQSMSPITSP-VESANQNNVMQPVATAPAPASAPLPQOSYPOP 206
Db 409 FETSSALSSSTIPLADLPPEPPPSVAISVPSAPAPPPPTLITTKKPSASSQSKISSSSSS 468
QY 207 FTTYSN-KAGMTSDXAMILLIOPITVASPPSPVAPPPPTSTGRASRKYVAPLAPSPAM 265
Db 469 AV 470
QY 266 EV 267

Search completed: Sun Sep 3 12:47:47 2000
Job time : 68 secs.

W093197 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit,
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:45:17 2000; MasPar time 32.19 Seconds

Tabular output not generated. 910.330 Million cell updates/sec

Title: >US-09-332-522B-6
Description: (1-1237) from US09332522B.pap
Perfect Score: 9226
Sequence: 1 MKNRKSKEFSKAMKFWPAC.....CDYGNELRVYRPIVADKCH 1237

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23586106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a:geneseq35
1:geneseqp

Statistics: Mean 40.573; Variance 189.529; scale 0.214

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	271	2.9	1170	1	W88447 yeast NPC1 protein ort	2.40e+11
2	244	2.5	587	1	Y03204 Amino acid sequence of	3.52e+09
3	237	2.5	1278	1	W88445 Human NPC1 (Niemann-Pi	1.27e+08
4	240	2.5	1296	1	W47157 Nevold basal cell carc	7.32e+09
5	240	2.5	1434	1	W52199 Mouse patched (ptc) pr	7.32e+09
6	240	2.5	1434	1	R94380 Mouse patched protein.	7.32e+09
7	240	2.5	1434	1	W72968 Mouse patched protein.	7.32e+09
8	240	2.5	1447	1	W52200 Human patched (ptc) pr	7.32e+09
9	240	2.5	1447	1	W72968 Human patched protein.	7.32e+09
10	240	2.5	1447	1	R75375 Human patched protein.	7.32e+09
11	218	2.4	1311	1	W72971 Precis coenia patched	3.95e+07
12	224	2.4	1319	1	W88446 Mouse NPC1 orthologue.	1.34e+07
13	215	2.3	1311	1	W52197 Precis coenia (buttef	6.76e+07
14	183	2.0	887	1	R58609 Hamster HMG-CoA reduct	1.91e+04
15	149	1.5	517	1	W85852 WD-40 domain-contg. be	6.02e+02
16	146	1.5	779	1	R85854 WD-40 domain-contg. CD	2.61e+01
17	128	1.4	318	1	R85855 WD-40 domain-contg. GB	1.76e+00
18	131	1.4	906	1	R85853 WD-40 domain-contg. be	1.10e+00
19	121	1.3	1054	1	R58610 Yeast HMG-CoA reductas	5.22e+01
20	116	1.3	1299	1	R86304 Drosophila patched pro	1.12e+01
21	120	1.3	1545	1	W53602 Rat sulphonylurea rece	6.09e+00
22	118	1.3	3588	1	R34712 Bacillus subtilis strA	8.26e+00
23	114	1.2	240	1	W41601 Secreted protein C195-	1.51e+01

24	111	1.2	251	1	W20277 H. pylori surface or m	2.37e+01
25	111	1.2	267	1	W21027 H. pylori putative mem	2.37e+01
26	108	1.2	325	1	R25074 Plap.	3.69e+01
27	108	1.2	325	1	R91347 Murine PLAP, for reduc	3.69e+01
28	108	1.2	325	1	R85872 WD-40 domain-contg. PL	3.69e+01
29	111	1.2	388	1	W86375 H. pylori GHPD 1138 pr	2.37e+01
30	114	1.2	417	1	W83388 Caenorhabditis elegans	1.51e+01
31	107	1.2	439	1	R85880 WD-40 domain-contg. YC	4.27e+01
32	109	1.2	514	1	R85881 WD-40 domain-contg. YC	3.18e+01
33	109	1.2	553	1	W10559 Barley gibberellin-reg	3.18e+01
34	110	1.2	704	1	W06079 Drosophila TATA-bindin	2.74e+01
35	110	1.2	704	1	R85877 WD-40 domain-contg. tr	2.74e+01
36	110	1.2	704	1	W25027 TATA-binding protein a	4.27e+01
37	107	1.2	704	1	R56495 TATA-binding protein-a	4.27e+01
38	107	1.2	704	1	W06085 Human TATA-binding pro	4.27e+01
39	115	1.2	798	1	R85879 WD-40 domain-contg. TU	1.30e+01
40	115	1.2	798	1	W33634 Yeast transcriptional	1.30e+01
41	115	1.2	798	1	R91304 TATA box binding prote	1.30e+01
42	115	1.2	798	1	R85883 WD-40 domain-contg. ye	1.30e+01
43	112	1.2	1045	1	R58611 Yeast HMG-CoA reductas	2.04e+01
44	108	1.2	1285	1	W72972 Drosophila melanogaste	3.69e+01
45	109	1.2	2629	1	W55885 Rat telomerase.	3.18e+01

ALIGNMENTS

Result 1
ID W88447 standard; Protein: 1170 AA.

AC W88447.
DT 26-APR-1999 (first entry)
DE Yeast NPC1 protein orthologue.
KW Niemann-Pick disease type C; NPC1 gene; yeast; orthologue;
KW diagnosis; therapy; animal model; cholesterol; neurodegeneration.
OS Saccharomyces cerevisiae.
PN W09901555-A1.
PD 14-JAN-1999.
PF 02-JUL-1998; U13862.
PR 03-JUL-1997; US-051682.
PA (USSH) US DEPT HEALTH & HUMAN RESOURCES.
PI Carstee ED, Gu J, Loftus SK, Morris JA, Pavan WJ,
PI Penchev PG, Rosenfield MA, Tagle DA;
DR WPI: 99-106056/09.
DR N-PSDB: X06875.
PT New isolated gene, NPC-1 - is associated with Niemann-Pick type C
PT disease, used to develop products for the study, diagnosis and
PT therapy of the disease.
PS Disclosure: Page 75-79; 101pp; English.
CC This polypeptide comprises the Saccharomyces cerevisiae orthologue
CC of the human NPC1 polypeptide (see W88445) that is associated with
CC Niemann-Pick disease type 2 (NP-C). The polypeptide shows
CC extensive identity (34%) and similarity (57%) to the human
CC NPC1 protein. Biochemical and genetic analysis of yeast, worm and
CC murine NPC1 model systems will provide resources for understanding
CC the role of NPC1 in intracellular cholesterol homeostasis and in
CC the aetiology of neurodegeneration in NP-C disease. The provision
CC of a human NPC1 cDNA sequence (see X06873) enables methods of
CC detecting the presence of mutations in the hNPC1 gene, and thereby
CC facilitates the determination of whether an individual is an NP-C
CC sufferer or carrier.
SQ Sequence 1170 AA;

Query Match 2.9%; Score 271; DB 1; Length 1170;

Best local similarity 21.2%; Pred. No. 2.40e+11;
Matches 36; Conservative 71; Mismatches 57; Indels 6; Gaps 5;

Db	557	DISTVAISYIMMFLVATYALRRKDG-KTRLLGSLGLIVLASIVCAAGFTLFGKSTL	615
Oy	305	ELVPYTVAFMFLVAVYPSVRKIDVRSRFLALCSVITTRGSLAMSLGCLCFPLTSL	364
Db	616	IIAEVIFPLLAIGDINFLITHEDRNCEOKPEPSIDOKIISAGRMSPLMSLQOT	675
Oy	365	OSKDIFPLVILVIGLNSLVITKSVSM-DET--FDVKIRVAQALSKGWHISKLTLEI	421


```

PR 16-MAY-1997: US-017906.
PR 17-MAY-1996: US-017905.
PR 21-MAY-1996: AU-000011.
PR 07-JUN-1996: AU-000363.
PR 14-JUN-1996: US-019765.
PA (USSH ) US SEC. DEPT HEALTH.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Chidambaram A, Christensen J, Dean MF, Gallani MR,
PI Gerard B, Gillies S, Goldstein A, Hahn H, Holmberg E,
PI Leifell DJ, Negus K, Pressman C, Shanley S, Smyth I,
PI Udden AB, Vorchevsky I, Mainwright B, Wicking C,
PI zaphiropoulos PG, Chevenix-trench, Totfgard R:
PI WPI: 98-008863/01.
DR N-PSDB: V15394.
PT Nevroid basal cell carcinoma syndrome tumour suppressor gene - useful
PT for detection of pre-disposition to basal cell carcinoma(s)
PS Disclosure: Fig 8: 148pp: English.
CC This is a nevroid basal cell carcinoma syndrome (NBCCS) (PTC) protein.
CC The encoding nucleic acid specifically hybridises, under stringent
CC conditions, to a second nucleic acid consisting of a 6568 (full-length
CC sequence), 1732 (exon 1a, b) (V15998) or 659 (exon 2a) (V15999) base pair
CC sequence. In the presence of a human genomic library. The PTC polypeptide
CC when presented as an antigen elicits the production of an antibody which
CC specifically binds to a polypeptide encoded by the above three sequences.
CC The NBCCS gene and its protein product, is a tumour suppressor, and is a
CC homologue of the Drosophila PATCHED (PTC) gene. Detection of the NBCCS
CC nucleic acid, in particular abnormal sequences, by hybridisation assays
CC is useful for detecting a predisposition to NBCCS or to a basal cell
CC carcinoma (also known as Gorlin syndrome). Alternatively, detection is of
CC the polypeptide and is carried out by immunoassay. Vectors comprising
CC this nucleic acid can be used to treat NBCCS. The PTC polypeptide can
CC mitigate symptoms of NBCCS in an organism. The NBCCS nucleic acid
CC includes one or more mutations, chosen from Exon-5 693insC, Exon-17
CC 2988del8bp, Exon-21 3538delG, Exon-22 64302T, Exon-12 1711insC, Exon-12
CC 1691nsA, Exon-16 2707delC, and Intron-17 3157-2A to G. The mutation may
CC be a nonsense or frameshift mutation. Frameshift mutations are chosen
CC from 244delC, 271nsA, 464insA, 693insC, 804del137, 877delG, 929delC,
CC 1370del176, 1393insrGCC, 1444del6, 1497dup8, 1639insA, 1711insC,
CC 1836delTC, 2320insnAA, 2392delA, 2574delA, 2583insC, 2596complex,
CC 2707delC, 2748insC, 2749dup7, 2988del8bp, 3014nsA, 3352delAT and
CC 3538delG. The mutation may be missense, chosen from G391T, G1148A,
CC G1368A, G1525T, C2050T, C2058T, C3015A, G3193C AND G4302T. Alternatively,
CC the mutation alters mRNA splicing and is chosen from A105>C, 3157-2A to
CC G and 1493-8ins12. All these mutations are claimed but their sequences
CC are not provided in the specification.
SO Sequence 1296 AA:

Query Match 2.68: Score 240, DB 1: Length 1296:
Best Local Similarity 22.8%: Pred. No. 7,32e-09:
Matches 38: Conservative 69: Mismatches 56: Indels 4: Gaps 3:

Db 268 VIRASGCTLMVACITMLRMDCSKSGQAVGLACVLVALSVAGCIGCISIGFVNA 347
Oy :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy 306 LVPYFAFMVAFVYFVSFKRIDVFESRPLCLACSVITTAGSLASIGCFEFTISLQ 365
Db 348 TTVCLPLFALGVCDVDFLLAHAFSETGONKRIPECDRTGCEKTRGASVALTSISNTA 407
Oy ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 366 SKQIFPLVILVIGLENSLVTR--SVYSNDETFDVKIRKQAOLSKEGWHSIKTILTEIT- 422
Db 408 FFMALIP-IPALRAFSLQAAYVVENFAMVLLIFPAILSMDLYRE 453
Oy ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 423 ILTIGLTFEVPVIOECIFAFVIGLLSDFMQLMFLFTLIANNIKRTE 469

RESULT
5
ID W52199 standard: Protein: 1434 AA.
AC W52199:
DT 25-JUN-1998 (first entry)
DE Mouse patched (Ptc) protein.
KW Patched protein; Ptc; cancer; tumour suppressor; cell adhesion promoter;
KW wound healing; ageing; mouse.
NN Mus sp.
NN W0974541-A2.

```

```

DB 04-DEC-1997: 009553.
PF 02-JUN-1997: 009553.
PR 31-MAY-1996: US-656055.
PA (REGC ) UNIV CALIFORNIA.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Epstein E, Goodrich LV, Johnson RL, Oro A, Scott MP:
DR WPI: 98-032648/03.
DR N-PSDB: V21389.
PT Patched protein other than Drosophila melanogaster patched protein -
PT used for characterizing the phenotype of a tumour
PS Claim 5: Pages 66-71: 86pp: English.
CC This is a mouse patched (ptc) protein. The encoding DNA can be used
CC to construct an expression cassette comprising an altered patch or
CC hedgehog gene. The expression cassette comprises a nucleic acid encoding
CC a patched protein other than a Drosophila melanogaster patched protein,
CC or fragment of at least 12 nucleotides in length, as other than an intact
CC chromosome under transcriptional control of a transcriptional initiation
CC region, and a transcriptional termination region, both functional in an
CC expression host. A genetically engineered mammalian cell comprising this
CC expression cassette as an extrachromosomal element or integrated into the
CC genome of the cell can be predisposed to develop basal cell carcinoma as
CC a result of the transfection. By analysing DNA, functional analysis of
CC patched protein function, or by detecting antibody binding to abnormal
CC patched protein, a genetic predisposition to developmental abnormalities
CC and cancer can be diagnosed. This analysis can also be used for
CC characterizing the phenotype of a tumour, particularly a carcinoma,
CC especially a basal cell carcinoma. The methods can also be used for
CC characterising transitional cell carcinoma of the bladder, meningiomas
CC medulloblastomas, etc. The modified cells comprising the expression
CC cassette can be used to determine the role of different exons of the
CC patched gene in oncogenesis, signal transduction, etc. Transgenic animal
CC models created from these cells can be used as animal models for
CC carcinomas of the skin. The patched protein of mosquito, butterfly or
CC beetle or alternatively, a mammalian patched protein of human or mouse
CC can be used to identify ligands or substrates that bind to, modulate, or
CC mimic the action of patched gene. These agents could be used as tumour
CC suppressors, cell adhesion promoters (e.g. in wound healing and ageing).
CC Sequence 1434 AA:
SQ
Query Match 2.6% Score 240; DB 1: Length 1434;
Best Local Similarity 22.8%; Pred. No. 7.32e-09;
Matches 38; Conservative 69; Mismatches 56; Indels 4; Gaps 3;
Db 425 VIRVAGSYLMLAACYCTMLRMKDCSKSQGAVGAGLVALLVAAGLIGLSTIGISFNAA 484
OY :::::||||| ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
306 LVPYVAMLVFAVYVSVRKIDVRSRPLFALCSVITTAGLSAMSLGCLFFPGLTISLQ 365
Db 485 TTVQLPFLAIGVGDVDFLLAHAFSETGQNKRIPEEDRTGECCKRGASVALTSISNVA 544
OY ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
366 SKDIFPYLVILVIGLNSLYTRK--SVVSMDETPEVVKIRVAQALSKEGMHSKLTLEIT- 422
Db 545 FFMAALIP-IALAPAFSLQAAVYVVFNFVAVLIFPAILSMDLYRRE 590
OY ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
423 ILTIGLATFVPIQEFCEFAIVGLISDFMQLMFLFSTILAMNKRTE 469
RESULT 6
ID R94380 standard; Protein: 1434 AA.
AC R94380.
DT 11-JUL-1996 (first entry)
DE Mouse patched protein.
KW Patched gene; PTC protein; embryo development; cellular regulation;
KW signal transduction; ligand; antibody; hedgehog protein.
OS Mus musculus.
PN W09611260-A1.
PD 18-APR-1996.
PF 06-OCT-1995: U13233.
PR 07-OCT-1994: US-319745.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Goodrich LV, Johnson RL, Scott MP:
DR WPI: 96-209842/21.
DR N-PSDB: 114218.
DNA encoding patched protein other than Drosophila patched protein -

```

PT used to produce antibodies which detect or inhibit patched protein
PT ligand signal transduction in cells
PS Disclosure: Page 46-51: 70pp; English.
CC Mouse patched protein (PTC) (R94380) was identified as the product
CC of a cDNA clone (T14218) derived from mouse limb bud. It has
CC about 38% identical amino acids to Drosophila PTC. PTC protein has
CC been proposed as a receptor for hedgehog protein on basis of genetic
CC experiments in flies. Murine PTC protein can be obt'd. in large
CC amounts by expression of the cDNA clone in transformed host cells.
CC It can be used to screen for agonists and antagonists, to isolate
CC its ligand, partic. Sonic hedgehog, to assay for the transcription
CC of ptc mRNA and to raise antibodies.
SQ Sequence 1434 AA;

Query Match 2.6%; Score 240; DB 1; Length 1434;

Best Local Similarity	22.8%;	Pred. No.	7.32e-09;
Matches	38;	Conservative	69;
		Mismatches	56;
		Indels	4;
		Gaps	3

Db 425 VIRVASYLLMLAYACTLMLRWDCKSGCAGVLGAVLVLAALVAAGLGCLIGISFNAA 484
:: : ::::| : : | : : | : | : ||| : :::
QY 306 LVPYVAFMLVEAYVFSVRKIDVFRRSPFLALCSVTITNGSLAMSLGLCEFFGLTSLQ 365

```

Db      485 TTQVLPFLALGVVDVFLLAHAFSETGONKRIPEDRTGCECLKRTGASVALTISNVTA 544
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     366 SKDIPYLVILVGLNSLVITK--SVSNDETFDVKIRVAQALSNEGWHISKTLTFEIT- 422

```

Db 545 FFMALIP-IPALRAFSIQAAVVVFNFAMVLLIFPAISMDLYRRE 590
:: : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Qy 423 ILTIGLAFVAVIQEFCIFAIVGLSDENQLMQLLESTILANIKRRE 469

RESULT	7
ID	W72968 standard; Protein; 1434 AA.

DI 26-JAN-1999 (first entry)
DE Mouse patched protein.
KW Mouse; patched gene; diagnosis; treatment; developmental disorder;

KM	sperm production; gene therapy.
OS	Mus sp.
PN	US5837538-A.

PR 07-OCT-1994; US-319745.
PR 06-OCT-1995; US-540406.
PR 06-OCT-1995; US-540406.
PR 07-OCT-1994; US-319745.

PA (SIND) UNIV DELAND SIMFORD JUNIOR.
PI Goodrich LV, Johnson RL, Scott MP;
DR WPI: 99-023461/02.
DR N-PSDB: V64092.

PT healing of injured tissue

CC The present sequence represents the mouse patched (*ptc*) protein. Cell
CC containing and expressing the *ptc* gene are used for the recombinant
CC production of the protein. These in turn are useful. (i) for generating
PS Claim 23; Column 49-56; 38pp; English.

CC antibodies (Ab); and (ii) to screen for specific-binding ligands
CC (potential therapeutic agonists and antagonists). The ptc gene, or its
CC fragments, are used to isolate related sequences from other mammals;
CC identify mutations (particularly those associated with genetic diseases

CC such as spina bifida and other developmental disorders).; to monitor
CC expression levels in testis (to determine relationship with sperm
CC production) and to isolate 5'-non-coding sequences (used to study
CC embryonic development and to provide regulated expression of proteins

CC The complete gene can be used in gene therapy, including expression c
CC antisense molecules, and to generate transgenic animals for studies c
CC embryonic development. Ab are used diagnostically to determine the
CC et al. protection cell surfaces and as competitive inhibitors of signal

CC transduction through the ptc ligand. Cells that have been engineered
CC express the ptc protein can be used to promote regrowth and healing c
CC damaged tissue (e.g. growth of new teeth) and regulation of the ptc

CC	Wnt-1 oncogene).
SQ	Sequence 1434 AA;

Query match	Score	DB 1;	Length
2.68;	240;	1;	1434;

Matches 38; Conservative 69; Mismatches 56; Indels 4; Gaps 3;

Db 425 VIRVASGYLLMLAYACTMTLRDCKSKSGAVGLAGVLLVALSVAAAGLGCLIGISTNNA 484
:: : :: : | : : : | : : ||| : ::
QY 306 LVPYVAFMLVEAYVYSVRKIDVERRSFELLALCSVITTAGSLANSLGCEFFGLTISLQ 365

Dd 485 TTQVLPFLALGVGDVDFLLAHAFSETGQNKRIPFEDRFGCECLKRTGTASVALTISINVT 544
 : :::: | :::: : : : : : : : : : :
Qy 366 SKDIPPYLIVLGLNSLVITK--SVASMDEFVDYKIRAAQLSNEGWHISKLTTLTELT- 422

Db 545 FEMALIP-IPALRAFSLQAAVVVVEFAMVLIFPAISMDLYRRE 590
:: : : : : : : : : : : : : : : :
QY 423 IITLGLATFPVVIQECIFAIVGLSDMLQMLLFTILAMNIKRT 469

RESULT	8
ID	W52200 standard; Protein; 1447 AA.

PC	M22200;
DT	25-JUN-1998 (first entry)
DE	Human patched (ptc) protein.
KW	patched protein; ptc; cancer; tumour suppressor; cell adhesion promoter;

AM wound healing, ageing, human.
OS Homo sapiens.
PN M09745541-A2.
PD 04-DEC-1997.

PR 02-JUN-1997, 000000.
PR 31-MAY-1996; US-656055.
PA (REGC) UNIV CALIFORNIA.
PA (STRD) UNIV LELAND STANFORD JUNIOR.

PT Patched protein other than Drosophila melanogaster patched protein -
DR N-PSDB: V21590.
F1 Epssteiner, GOODRICH LV, JOHNSON RV, OLO R, SCOLL MF,
WPI: 98-032648/03.

CC This is a human patched (ptc) protein. The encoding DNA can be used
to construct an expression cassette comprising an altered patch or
CC to construct an expression cassette comprising an altered patch or
PS Claim 4; Pages 76-80; 86pp; English.
P1 used to characterising the phenotype of a mutant

CC chromosome under transcriptional control of a transcriptional initiation
CC or fragment of at least 12 nucleotides in length, as other than an intact
CC patched protein other than a *Drosophila melanogaster* patched protein,
CC neogenog gene. The expression cassette comprises a nucleic acid encoding

CC region, and a transcriptional termination region, both functional in an
CC expression host. A genetically engineered mammalian cell comprising this
CC expression cassette as an extrachromosomal element or integrated into the
CC genome of the cell can be predisposed to develop basal cell carcinoma as

CC a result of the transfection. By analysing DNA functional analysis of CC patched protein function, or by detecting antibody binding to abnormal CC patched protein, a genetic predisposition to developmental abnormalities CC and cancer can be diagnosed. This analysis can also be used for

CC characterising the phenotype of a tumour, particularly a carcinoma, CC especially a basal cell carcinoma. The methods can also be used for CC characterising transitional cell carcinoma of the bladder, meningiomas

CC cassette can be used to determine the role of different exons of the
CC patched gene in oncogenesis, signal transduction, etc. Transgenic animal
CC models created from these cells can be used as animal models for

CC beetle or alternatively, a mammalian patched protein of human or mouse
CC can be used to identify ligands or substrates that bind to, modulate, or
CC mimic the action of patched gene. These agents could be used as tumour

SQ	Sequence	1447 AA;
Query Match	2.68;	Score 240; DB 1; Length 1447;

Db 439 VIRVASYLLMLAYACTMLPMDCKSQGAVGLAGVLLVALSVAGGLGCLSLIGISFNAA 498
Matches 38; Conservative 69; Mismatches 56; Indels 4; Gaps 3;

[illegible]

RESULT 9
ID W72969 standard; Protein; 1447 AA.
AC W72969;
DT 26-JAN-1999 (first entry)
DE Human patched protein.
KW Human; patched gene; diagnosis; treatment; developmental disorder;
KW cancer; healing; injured tissue; ptc; spina bifida; Mnt-1 oncogene;
KW sperm production; gene therapy.
OS Homo sapiens.
PN US837538-A.
PD 17-NOV-1998.
PF 06-OCT-1995: 540405.
PR 06-OCT-1995: US-540405.
PR 07-OCT-1994: US-319745.
PA (SHRD) UNIV LELAND STANFORD JUNIOR.
PI Goodrich LV, Johnson RL, Scott MP;
DR WPI: 99-023461/02.
DR N-PSDS: V64093.
PT Nucleic acid encoding vertebrate patched protein and related
PT transformants - used to express poly(peptide)s , useful for
PT diagnosis and treatment of developmental disorders of cancer, and in
PT healing of injured tissue
PS Claim 23: Column 63-70: 38pp; English.
CC The present sequence represents the human patched (ptc) protein. Cells
CC containing and expressing the ptc gene are used for the recombinant
CC production of the protein. These in turn are useful: (1) for generating
CC antibodies (Ab); and (1i) to screen for specific-binding ligands
CC (potential therapeutic agonists and antagonists). The ptc gene, or its
CC fragments, are used to isolate related sequences from other mammals; to
CC identify mutations (particularly those associated with genetic diseases
CC such as spina bifida and other developmental disorders); to monitor
CC expression levels in tests (to determine relationship with sperm
CC production) and to isolate 5'-non-coding sequences (used to study
CC embryonic development and to provide regulated expression of proteins).
CC The complete gene can be used in gene therapy, including expression of
CC antisense molecules, and to generate transgenic animals for studies of
CC embryonic development. Ab are used diagnostically to determine the
CC ptc protein on cell surfaces and as competitive inhibitors of signal
CC transduction through the ptc ligand. Cells that have been engineered to
CC express the ptc protein can be used to promote regrowth and healing of
CC damaged tissue (e.g. growth of new teeth) and regulation of the ptc
CC protein expression may be useful in cancer treatment (it may control the
CC Mnt-1 oncogene).
CC Sequence 1447 AA:

Query Match:	2.6%	Score 240;	DB 1;	Length 1447;
Best Local Similarity:	22.8%;	Pred. No. 7,32e-09;		
Matches	38;	Conservative	69;	Mismatches 56; Indels 4; Gaps 3;
Db	439	VIRVASGVLMLAVACTMLRMDCKSGOANGVLGLVLLVASVAGLGCLSLGISFNMA	498	
Oy	306	LVPITVAFMLVFAYVTSVKRIDVFRSRFLIADCSTTTGSLMSIGLCPFGELTISLO	365	
Db	499	TTQVLPFLIAGVGDVDFLLAHAFSEFGONKRIPEFEDTECELTARTASVALTISINVA	558	
Oy	366	SKDIFPVILVILVGLNSLVIIR--SVASMDETPOVKIRVAQALSKBGMHLSKLTLEIT-	422	
Db	559	FFMAALID-IPALRAFSLQAADVVENFAMVLLIFPAILSLSDLYRRE	604	
Oy	423	ILITGLAFVPIVIOEFCFAIVGLSLSPLOMLFSTLIANNIRTE	469	

RESULT 10
ID R75375 standard; Protein; 1447 AA.
AC R75375;
DT 30-JUL-1996 (first entry)
DE Human patched protein.
KW Patched gene; PTC protein; embryo development; cellular regulation;
KW signal transduction; ligand; antibody; hedgehog protein.
OS Homo sapiens.
PN M09611260-A1.
PD 18-APR-1996.
PF 06-OCT-1995; U13233.
PR 07-OCT-1994; US-319745.
PA (STRD.) UNIV LEIAND STANFORD JUNIOR.
PI Goodrich LV, Johnson RL, Scott MP;
WP1: 96-2098442/21.
DR N-PSDB: T14220.
PT DNA encoding patched protein other than Drosophila patched protein
PT used to produce antibodies which detect or inhibit patched protein
PT ligand signal transduction in cells
PS Example: Page 56-61; 70pp; English.
CC Human patched protein (PTC) (R75375) was identified as the product
CC of a cDNA clone (T14220) derived from human lung. It has 968
CC identity and 96% similarity to mouse PTC. PTC protein has
CC been proposed as a receptor for hedgehog protein on the basis of
CC genetic experiments in flies. Human PTC protein can be obtd. in
CC large amounts by expression of the cDNA clone in transformed host
CC cells. It can be used to screen for agonists and antagonists, to
CC isolate its ligand, partic. Sonic hedgehog, to assay for the
CC transcription of ptc mRNA and to raise antibodies.
SQ Sequence 1447 AA;

[illegible]

RESULT	11	
ID	W72971	standard; Protein: 1311 AA.
AC	W72971;	
DT	26-JAN-1999	(first entry)
DE	Precis coenla patched amino acid sequence.	
KW	cancer; gene; ptc; diagnosis; treatment; developmental disorder	
KW	cancer; healing; injured tissue; spina bifida; Mnt-1 oncogene;	
KW	sperm production; gene therapy.	
OS	Precis coenla.	
FH	Key	Location/Qualifiers
FT	Misc_difference	348
FT	/note=	"unspecified"
FT	Misc_difference	908
FT	/note=	"unspecified"
FN	US5837538-A.	
PD	17-NOV-1998.	
PE	06-OCT-1995; 540406.	
PR	06-OCT-1995; US-540406.	
PR	07-OCT-1994; US-319745.	
PA	(STRD) UNIV LRELAND SPANFORD JUNIOR.	
PI	Goodrich LV Johnson RL, Scott MP;	
PI	WPI: 99-023461/02.	
DR		
CT	Nucleic acid encoding vertebrate patched protein and related	
CT	transformants - used to express poly(peptide(s)) useful for	

Query Match	2.4%;	Score 218;	DB 1;	Length 1311;
Best Local Similarity	23.18;	Pred. No. 3.95e-07;		
Matches	36;	Conservative	60;	Mismatches 57;
			Indels	3;
			Gaps	3;

[illegible]

RESULT 12
ID W88446 standard; Protein; 1319 AA.
AC W88446;
DT 26-APR-1999 (first entry)
DE Mouse NPC1 orthologue.
KW Niemann-Pick disease type C; NPC1 gene; mouse; orthologue;
KW diagnosis; therapy; cholesterol; neurodegeneration.
OS Mus sp.
PN W09901555-A1.
PD 14-JAN-1999.
PF 02-JUL-1998; U13862.
PR 03-JUL-1997; US-051682.
PA (USSH) US DEPT HEALTH & HUMAN RESOURCES.
PI Carstee ED, Gu J, Loftus SR, Morris JA, Pavan WJ,
PI Pechter PG, Rosenfeld MA, Tagle DA;
DR WPI: 99-106056/09.
DR N-PSDB; X06874.
PT New isolated gene, NPC-1 - is associated with Niemann-Pick type C
PT disease, used to develop products for the study, diagnosis and
PT therapy of the disease
PS Claim 1; Page 66-70: 101pp; English.
CC This polypeptide comprises the murine orthologue of human NPC-1 (see
CC W99445) that is associated with Niemann-Pick disease type C (NP-C).
CC The amino acid sequence, predicted from an isolated cDNA clone
CC (see W06874), comprises a putative N-terminal signal peptide
CC followed by a domain that is unique to the NPC1 orthologues (in
CC mouse, human, Caenorhabditis elegans and Saccharomyces cerevisiae,
CC see W88445-48), and 13 putative transmembrane domains that include
CC a potential sterol-sensing domain. Biochemical and genetic
CC analysis of NPC1 using mouse model systems will provide resources

CC for understanding the role of NPC1 in intracellular cholesterol
CC homeostasis and in the etiology of neurodegeneration in NP-C
CC disease. The provision of a human NPC1 cDNA sequence (see X06873)
CC provides methods of detecting the presence of mutations in the
CC NPC1 gene, and thereby facilitates the determination of whether an
CC individual is an NP-C sufferer or carrier.
SQ Sequence 1319 AA:

Query Match	2.4%;	Score 224;	DB 1;	Length 1319;
Best Local	Similarity 26.3%;	Pred. No. 1.34e-07;		
Matches	45;	Conservative	61;	Mismatches 54; Indels 11; Gaps 9;

Dd 667 IISVAVMEFLIYSLALGHIOSSKLLV-DSKISUGIAGILIVLESSVACSGIFSYMGNPLT 725
::
Qy 306 LVPVTAFMELV-FAYVF-SVRKDIVERFRLALCSVITTAGLSAMSTGICEFFGLITS 363S
::
Dd 726 LIVIEVIPLELVANGVNINFILOTY-QREBRDOEF-LDOQLGRILIG-EVAFTMLFSF 782
- ::
Qy 364 LOSKDIPFYEVILEGENSVLITKSVMDETFVKIRVAOALSKE-GWHSIKTL-LT-- 419

Db 783 SETSAEFFGALLSMFAHTPSLFMGMAVLDFLIOLICFPASLLGDIDKKR 833
Qy 420 -EIIILIGATLEVPIOECFIAVBLEDDEMLOMLEFTLIAANMKRR 469

RESULT 13
ID W52197 standard; Protein; 1311 AA.
AC W52197;
DT 25-JUN-1998 (first entry)
DE Precis coenia (butterfly) patched (ptc) protein.
KW patched protein; ptc; cancer; tumour suppressor; cell adhesion promoter;
KW wound healing; ageing; Precis coenia; butterfly.
OS Precis coenia.
PN M09745541-A2.
PD 04-DEC-1997.
PF 02-JUN-1997; U09553.
PR 31-MAY-1996; U5-656055.
PA (REGC) UNIV CALIFORNIA.
PA (STRD) UNIV IELAND STANFORD JUNIOR.
PI Epstein E, Goodrich LV, Johnson RL, Oro A, Scott MP;
OR WPI: 98-032648/03.

PT Patched protein other than Drosophila melanogaster patched protein -
Pr used for characterizing the phenotype of a tumour
PS Claim 2, Pages 52-55; 86pp; English.
CC This is a Precis coena (butterfly) patched (ptc) protein and the
CC encoding DNA can be used to construct an expression cassette comprising
CC an altered patch or hedgehog gene. The expression cassette comprises a
CC nucleic acid encoding a patched protein other than a Drosophila
CC melanogaster patched protein, or fragment of at least 12 nucleotides in
CC length, as other than an intact chromosome under transcriptional control
CC of a transcriptional initiation region, and a transcriptional termination
CC region, both functional in an expression host. A genetically engineered
CC mammalian cell comprising this expression cassette as an extrachromosomal
CC element or integrated into the genome of the cell can be predisposed to
CC develop basal cell carcinoma as a result of the transfection. By
CC analysing DNA, functional analysis of patched protein function, or by
CC detecting antibody binding to abnormal patched protein, a genetic
CC predisposition to developmental abnormalities and cancer can be
CC diagnosed. This analysis can also be used for characterising the
CC phenotype of a tumour, particularly a carcinoma, especially a basal cell
CC carcinoma. The methods can also be used for characterising transitional
CC cell carcinoma of the bladder, meningiomas medulloblastomas, etc. The
CC modified cells comprising the expression cassette can be used to
CC determine the role of different exons of the patched gene in oncogenesis,
CC signal transduction, etc. Transgenic animal models created from these
CC cells can be used as animal models for carcinomas of the skin. The
CC patched protein of mosquito, butterfly or beetle or alternatively, a
CC mammalian patched protein of human or mouse can be used to identify
CC ligands or substrates that bind to, modulate, or mimic the action of
CC patched gene. These agents could be used as tumour suppressors, cell
CC adhesion promoters (e.g. in wound healing and ageing).
SQ Sequence 1311 AA;

THIS PAGE BLANK (USPTO)

NWSESE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:43:14 2000; MasPar time 58.87 Seconds
Tabular output not generated. 991.194 Million cell updates/sec

Title: >US-09-332-522B-6
Description: (1-1237) from US09332522B.pep
Perfect Score: 9226
Sequence: 1 MNKTKSKHFKAMKFWPAC.....CDYGNFIRVRFPIVADKCH 1237

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir64
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 55.949; Variance 117.221; scale 0.477

Pred: No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	271	2.9	1170	2	S52525	probable membrane pro 1.47e-23
2	223	2.4	1182	2	T13952	membrane protein pich 9.23e-16
3	212	2.3	932	1	A31898	hydroxymethylglutaryl 4.87e-14
4	210	2.3	1055	2	T05663	hypothetical protein 9.95e-14
5	183	2.0	887	1	RDHYE	hydroxymethylglutaryl 1.20e-09
6	183	2.0	887	1	A23586	hydroxymethylglutaryl 1.20e-09
7	183	2.0	888	1	RDHYE	hydroxymethylglutaryl 1.20e-09
8	178	1.9	856	1	S30338	hydroxymethylglutaryl 6.49e-09
9	178	1.9	856	1	A35728	hydroxymethylglutaryl 6.49e-09
10	177	1.9	916	2	S32572	hydroxymethylglutaryl 9.07e-09
11	177	1.9	1053	2	S72194	hydroxymethylglutaryl 1.58e-05
12	154	1.7	293	2	B71215	hypothetical protein 1.58e-05
13	155	1.7	701	2	T16607	hypothetical protein 1.15e-05
14	145	1.6	257	2	S62507	hypothetical protein 2.55e-04
15	150	1.6	308	2	E75029	protein-export membra 5.50e-05
16	147	1.6	431	2	S49821	beta-transducin repea 7.49e-05
17	149	1.6	518	2	B48088	hypothetical protein 2.55e-04
18	145	1.6	1015	2	T15830	myosin-heavy-chain ki 1.88e-04
19	145	1.6	1146	2	A55332	probable translocator 3.74e-03
20	136	1.5	425	2	S78258	hypothetical protein 4.67e-04
21	143	1.5	495	2	T04784	conserved hypothetical 1.15e-03
22	140	1.5	750	1	D69403	cell division control 2.79e-03
23	137	1.5	779	2	S56245	

24	134	1.5	888	2	E71280	probable antibiotic t 6.70e-03
25	129	1.4	282	2	D64456	protein-export membra 2.81e-02
26	130	1.4	290	2	T02300	GTP-binding regulator 2.11e-02
27	129	1.4	308	2	S74813	hypothetical protein 2.81e-02
28	128	1.4	318	2	S11904	GTP-binding regulator 3.73e-02
29	128	1.4	325	2	T06784	GTP-binding protein b 3.73e-02
30	126	1.4	325	2	T09613	probable GTP-binding 6.52e-02
31	126	1.4	334	2	T03764	conserved hypothetical 4.93e-02
32	127	1.4	428	2	B61817	PLI protein - Arabid 1.59e-02
33	131	1.4	486	2	S49820	hypothetical protein 8.95e-03
34	133	1.4	714	2	T16126	probable membrane pro 2.11e-02
35	130	1.4	817	2	S51445	coatomer complex beta 1.59e-02
36	131	1.4	906	2	S53312	Golgi-associated part 8.62e-02
37	131	1.4	906	2	S35342	H+-transporting Atpas 8.62e-02
38	125	1.4	952	2	S3548	H+-transporting Atpas 8.62e-02
39	125	1.4	953	2	T12577	probable protein 1.14e-01
40	125	1.4	1253	2	S62544	protein-export membra 1.14e-01
41	124	1.3	257	2	B69214	spermidine/putrescine 1.14e-01
42	124	1.3	286	2	S73424	guanine nucleotide re 1.14e-01
43	123	1.3	327	2	S48839	probable rodA protein 1.14e-01
44	124	1.3	469	2	G70699	hypothetical protein 1.14e-01
45	124	1.3	614	2	S58306	

ALIGNMENTS

RESULT ENTRY TITLE	1	ALIGNMENTS
S52525	#type complete	
probable membrane protein YPL006w - yeast (Saccharomyces cerevisiae)		
ALTERNATE_NAMES	hypothetical protein YPL11w: hypothetical protein YP8132.07	
ORGANISM	#formal_name Saccharomyces cerevisiae	
DATE	08-May-1995 #sequence_revision 21-Jul-1995 #text_Change 06-Feb-1998	
ACCESSIONS	S52525; S59687	
REFERENCE	S52519	
#authors	Badcock, K.; Churcher, C.	
#submission	submitted to the EMBL Data Library, February 1995	
#accession	S52525	
##molecule_type DNA		
##residues	1-1170 ##label BAD	
##cross-references EMBL:248483; NID:9683777; PID:9683784; MIPS:YPL006w		
##experimental_source strain AB972		
REFERENCE	S59677	
#authors	Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.; Wang, Y.; Winnett, E.	
#submission	submitted to the EMBL Data Library, August 1995	
#description	The sequence of Saccharomyces cerevisiae chromosome XVI left arm.	
#accession	S59687	
##molecule_type DNA		
##residues	1-1170 ##label HAL	
##cross-references EMBL:033335; NID:965076; PID:965087; MIPS:YPL006w		
GENETICS		
#gene	SGD:NCRL	
##cross-references SGD:S0005927; MIPS:YPL006w		
KEYWORDS	#map_position 16L	
FEATURE	transmembrane protein	
1-17		
258-274	#domain transmembrane #status predicted #label TM1	
341-357	#domain transmembrane #status predicted #label TM2	
585-601	#domain transmembrane #status predicted #label TM3	
614-630	#domain transmembrane #status predicted #label TM4	
667-683	#domain transmembrane #status predicted #label TM5	
698-714	#domain transmembrane #status predicted #label TM6	
751-767	#domain transmembrane #status predicted #label TM7	
1004-1020	#domain transmembrane #status predicted #label TM8	
1027-1043	#domain transmembrane #status predicted #label TM9	
1051-1067	#domain transmembrane #status predicted #label TM10	
1103-1119	#domain transmembrane #status predicted #label TM11	
1137-1153	#domain transmembrane #status predicted #label TM12	
SUMMARY	#length 1170	
	#molecular_weight 132644	
	#checksum 5191	

```
#cross-references MIMD:89054023          urchin embryo. Deduced structure and regulatory properties
#accession      A31898
##molecule_type mRNA
##residues      1-932 ##label WOO
##cross-references GB:J04200; MID:g161522; PIDN:AAA30060.1; PID:g161523
##note           the authors rearranged portions of the coding region in
                Figure 2, and the above sequence is taken directly
                from Figure 3: it matches the translation of the
                nucleotide sequence that the author submitted to
                Genbank
            #note           the authors translated the codon GGA for residue 805 as
                Glu
REFERENCE      A28367
#authors       Woodward, H.D.; Allen, J.M.C.; Lennarz, W.J.
#journal       J. Biol. Chem. (1988) 263:2513-2517
#title         3-hydroxy-3-methylglutaryl coenzyme A reductase in the sea
                urchin embryo is developmentally regulated.
#cross-references MIMD:8815403
#accession     A28367
##molecule_type mRNA
##residues     689-735 ##label WO2
COMMENT        This transmembrane glycoprotein of the endoplasmic reticulum is
CLASSIFICATION Involved in the control of cholesterol biosynthesis.
KEYWORDS       superfamily hydroxymethylglutaryl-CoA reductase (NADPH)
                cholesterol biosynthisis; endoplasmic reticulum;
                glycoprotein; NADP; oxidoreductase
FEATURE
279,850,886,930 #binding site carbohydrate (Asn) (covalent) #status
                  predicted
SUMMARY         #length 932 #molecular-weight 100965 #checksum 3584
Query Match    2.3%; Score 212; DB 1; Length 932;
Best Local Similarity 24.0%; Pred. No. 4,87e-14;
Matches 36; Conservative 58; Mismatches 53; Indels 3; Gaps 3;
Db 72 LAVAYHILQIFKRLTKTSSKIILGAGLFTFFSFLFSANVHLFGLELTGLNEALPFLL 130
OY 315 LVFAYVFYSVKRIDFVSREFLLACSVTTAGSIAMSLGICFEFGTLTSLSQSDIPPLYV 374
Db 131 LLIDLTASALTFRPALSTSTONEVDVIARGMALIGPTITLDIVVTTLV-SIGTMSSIR 189
OY 375 ILVGLENSIVTKSVSMDETFPDVKIKVAQALSKEGNHIS-KLLDLREITILLTGIANFPV 433
Db 190 KMEVPCCGEIIISLIANTFVEFTFPACLST 219
OY 434 VIOECIFAIVIGLSDFMLOMLSTILIAM 463
RESULT 4
ENTRY T05663 #type complete
TITLE hypothetical protein F22I13.120 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
                cress
DATE 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
                23-Jul-1999
ACCESSIONS T05663
REFERENCE Z15420
#authors Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft,
                I.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
#submission submitted to the Protein Sequence Database, February 1999
#accession T05663
##molecule_type DNA
##residues 1-1055 ##label BEV
#cross-references EMBL:AL035339
#experimental_source cultivar Columbia; BAC clone F22I13
GENETICS
#map_position 4
#introns 24/3: 60/3: 99/3: 150/1: 193/1: 216/3: 278/3: 297/2: 336/3:
                364/3: 396/3: 419/3: 511/3: 548/3: 577/3: 620/3: 667/1:
                735/1: 747/3: 775/3: 825/3: 874/1: 895/3: 927/1: 958/3:
                961/2: 1003/3: 1026/3
F22I13.120
```

```

SUMMARY      #length 1055  #molecular-weight 116148  #checksum 6862

Query Match      2.3%:  Score 210;  DB 2;  Length 1055;
Best Local Similarity 24.1%:  Prod. No. 9,95e-14;
Matches 41;  Conservative 56;  Mismatches 66;  Indels 7;  Gaps 5;

Db      544  DVITTAASYLVMEFVISVTLGDAPQEFYFYSKRLGSLGVLLVSLVGSVGV-PSAL 602
      305  ELVPYVAFMLVFAV--VYFSS-VRKIDVF--RSRFLALCSVITTAGSLAMSLGICFF-F 358
      603  GKRSITLIMEVIFPLVLAVGVDMCLIVAAVKQKPRVLSLEQRISSALVVEGSPITIASL 662
      339  GLTISLQSDKIDPFYLVILVGLNSLVITKSVAMDETFDKIRVAQALSKRGWHISKTL 418
      663  SEVLAFAVGAFFVMPACRIFFSMFALAIMDFLOITAFVALLVFDCKRS 712
      419  TITITITIGLAFVFPVIOEFCIFAIYGLSLDFMLQMLLESTILAMNKRT 468

RESULT 5
ENTRY   RDHYE      #type complete
TITLE   hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34) -
        Chinese hamster
ORGANISM #formal_name Citicetusulus griseus #common_name Chinese hamster
DATE     15-Nov-1984 #sequence_revision 15-Nov-1984 #text_change
        24-Sep-1999
ACCESSIONS A93328: A00357
REFERENCE  A93328
          #authors
            Chin, D.J.; Gil, G.; Russell, D.W.; Liscum, L.; Luskey, K.L.;
            Basu, S.K.; Okayama, H.; Berg, P.; Goldstein, J.L.; Brown,
            M.S.
          #journal
            Nature (1984) 308:613-617
          #title
            Nucleotide sequence of 3-hydroxy-3-methyl-glutaryl coenzyme A
            reductase, a glycoprotein of endoplasmic reticulum.
          #cross-references MUID:84168178
          #accession
            A93328
          #molecule_type mRNA
          #residues
            1-887 ##label CHI
          #cross-references GB:L00183; GB:X00494; NID:g191109; PIDW:AAA36989.1;
            PID:g387052
          ##note
            673-Cys was also found

REFERENCE  A90852
          #authors
            Reynolds, G.A.; Basu, S.K.; Osborne, T.F.; Chin, D.J.; Gil,
            G.; Brown, M.S.; Goldstein, J.L.; Luskey, K.L.
          #journal
            Cell (1984) 38:275-285
          #title
            HMG CoA reductase: a negatively regulated gene with unusual
            promoter and 5' untranslated regions.
          #cross-references MUID:84282711
          #contents
            annotation: introns
          #reference
            A92521
          #authors
            liscum, L.; Finer-Moore, J.; Stroud, R.M.; Luskey, K.L.;
            Brown, M.S.; Goldstein, J.L.
          #journal
            J Biol. Chem. (1985) 260:522-530
          #title
            Domain structure of 3-hydroxy-3-methylglutaryl coenzyme A
            reductase, a glycoprotein of the endoplasmic reticulum.
          #cross-references MUID:85080129
          #contents
            annotation: domains
          #COMMENT
            This transmembrane glycoprotein of the endoplasmic reticulum is an
            enzyme of cholesterol biosynthesis. An enzymatically active 62k
            fragment can be released from the membrane by cleavage (between
            residues 368-380) by an endogenous protease; this fragment can be
            further cleaved (between residues 450-470) by exogenous protease,
            producing a 53k fragment, which retains full enzymatic activity.

GENETICS
#introns
55/3: 93/1: 122/2: 150/3: 186/1: 221/3: 260/1: 314/2: 397/1:
455/3: 520/3: 573/3: 626/2: 661/3: 718/3: 765/3: 818/3:
870/2
#superfamily hydroxymethylglutaryl-CoA reductase (NADPH)
#cholesterol biosynthesis: endoplasmic reticulum:
#glycoprotein: oxidoreductase: transmembrane protein

FEATURE
1-339
10-39
#domain membrane-bound #label MBD\
#domain transmembrane #status predicted #label TM1\

```

Query Match	2.0%;	Score 183;	DB 1;	Length 887;
Best Local Similarity 18.7%;	Pred. No. 1,20e-09;			
Matches 31;	Conservative 68;	Mismatches 62;	Indels 5;	Gaps 5;
57-78	#domain transmembrane	#status predicted	#label TM2\	
92-114	#domain transmembrane	#status predicted	#label TM3\	
124-149	#domain transmembrane	#status predicted	#label TM4\	
160-187	#domain transmembrane	#status predicted	#label TM5\	
192-220	#domain transmembrane	#status predicted	#label TM6\	
315-339	#domain transmembrane	#status predicted	#label TM7\	
400-887	#domain catalytic hydrophilic	#label HYL\		
281	#binding site carbohydrate (Asn)	(covalent)	#status predicted	
SUMMARY	#length 887	#molecular-weight 97080	#checksum 872	
Query Match	2.0%;	Score 183;	DB 1;	Length 887;
Best Local Similarity 18.7%;	Pred. No. 1,20e-09;			
Matches 31;	Conservative 68;	Mismatches 62;	Indels 5;	Gaps 5;
61	DIILITTRCALIYIYFOPONRLQSGKYLIGAGLFTIFSSFPVS-TVIYHF-LDKEL	118		
305	ELVRYTAAFLMVFYVSVRKIDVRSRFLALCSVITTAAGSLMSGLGCFPGGLTSL	364		
119	TGLNEALPEFLLLIDISRASALAKFLSSNODDEVRENTARGMALIGPTFLDALVECLY	178		
365	QS-KDIPYIVLVGLGSLVITKRSVSDDEFEDVKIRAKOALSSEGHHSITLLTLEITI	423		
179	IGVGTMSGVRQLEIMCCFGMSVLANYPFMTFFPACVSLVELSR	224		
424	LTIGLAFVPIYIOEFCIFALVGLSDPMLQMLLEST-1-LAMNIKR	467		
RESULT	6	A23586	#type complete	
ENTRY	hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) - golden			
TITLE	hamster			
ORGANISM	#formal_name Mesocricetus auratus #common_name golden hamster			
DATE	10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change			
ACCESSIONS	A23586			
REFERENCE	A23586			
#authors	Skalniak, D.G.; Simoni, R.D.			
#journal	DNA (1985) 4:439-444			
#title	The nucleotide sequence of Syrian hamster HMG-CoA reductase cDNA.			
#cross-references	MJ01:86135263			
#accession	A23586			
#molecule_type	mRNA			
#residues	1-887 #label SKA			
CLASSIFICATION	#superfamily hydroxymethylglutaryl-CoA reductase (NADPH)			
KEYWORDS	oxidoreductase; transmembrane protein			
SUMMARY	#length 887 #molecular-weight 96970 #checksum 74			
Query Match	2.0%;	Score 183;	DB 1;	Length 887;
Best Local Similarity 18.7%;	Pred. No. 1,20e-09;			
Matches 31;	Conservative 68;	Mismatches 62;	Indels 5;	Gaps 5;
61	DIILITTRCALIYIYFOPONRLQSGKYLIGAGLFTIFSSFPVS-TVIYHF-LDKEL	118		
305	ELVRYTAAFLMVFYVSVRKIDVRSRFLALCSVITTAAGSLMSGLGCFPGGLTSL	364		
119	TGLNEALPEFLLLIDISRASALAKFLSSNODDEVRENTARGMALIGPTFLDALVECLY	178		
365	QS-KDIPYIVLVGLGSLVITKRSVSDDEFEDVKIRAKOALSSEGHHSITLLTLEITI	423		
179	IGVGTMSGVRQLEIMCCFGMSVLANYPFMTFFPACVSLVELSR	224		
424	LTIGLAFVPIYIOEFCIFALVGLSDPMLQMLLEST-1-LAMNIKR	467		
RESULT	7	RDHDE	#type complete	
ENTRY	hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34) -			
TITLE	human			
ORGANISM	#formal_name Homo sapiens #common_name man			
DATE	04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change			
	11-Jun-1999			

ACCESSIONS A00356
 REFERENCE A00356
 #authors Luskey, K.L.; Stevens, B.
 #journal J. Biol. Chem. (1985) 260:10271-10277
 #title Human 3-hydroxy-3-methylglutaryl coenzyme A reductase. Conserved domains responsible for catalytic activity and sterol-regulated degradation.
 #cross-references MIM:85261451
 #accession A00356
 #molecule_type mRNA
 #residues 1-888 #label LUS
 #cross-references GB:M11058; NID:g184243; PIDN:AAA52679.1; PID:g306865
 COMMENT This transmembrane glycoprotein of the endoplasmic reticulum is involved in the control of cholesterol biosynthesis.
 GENETICS
 #gene GDB:HMGR
 #cross-references GDB:119312; OMIM:142910
 #map_position 5q13.3-5q14
 CLASSIFICATION #superfamily hydroxymethylglutaryl-CoA reductase (NADPH)
 KEYWORDS cholesterol biosynthesis; endoplasmic reticulum; glycoprotein; membrane protein; oxidoreductase
 FEATURE
 1-339 #domain membrane-bound #label MBD
 340-449 #domain linker #label LKR
 450-888 #domain catalytic #label CAT
 281,518,870 #binding_site carbohydrate (Asn) (covalent) #status predicted
 SUMMARY #length 888 #molecular-weight 97475 #checksum 2950
 Query Match 2.08; Score 183; DB 1; Length 888;
 Best Local Similarity 18.7%; Pred. No. 1.20e-09;
 Matches 31; Conservative 68; Mismatches 62; Indels 5; Gaps 5;
 Db 61 DIILITRICALIYTFQFONLRQSGKIIGAGLFTTSSFFVS-TVVIHF-LDKEL 118
 QY 305 ELVPTVAVMLFAVYVFSRKRIDVRSRFLALCSVITTAGSLASGLCFEFGITLSL 364
 Db 119 TGLNEALPFFLLIDLSRASLAKFALSSNSODEVRENIRAGMAILGPTTEALVECV 178
 QY 365 QS-KDIFPVLIVLGLNSLVITKSVSMDETFDKIRVAQALSKGWHISKTLLEITI 423
 Db 179 IGVGTMSGVROLEIMCCFGCMSVLANYFAFMTFFPACVSLVELSR 224
 QY 424 LITGLATFVPVIOEFCIFAIYGLSDFMQLFST-I-LAMNIKR 467
 RESULT 8
 ENTRY S30338 #type complete
 TITLE hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34) - German cockroach
 ORGANISM #formal_name Blattella germanica #common_name German cockroach
 DATE 02-Dec-1993 #sequence_revision 03-Aug-1995 #text_change 21-Nov-1998
 ACCESSIONS S30338
 REFERENCE S30338
 #authors Martinez-Gonzalez, J.; Buesa, C.; Paulachs, M.D.; Belles, X.; Hegardt, F.G.
 #journal Eur. J. Biochem. (1993) 213:233-241
 #title Molecular cloning, developmental pattern and tissue expression of 3-hydroxy-3-methylglutaryl coenzyme A reductase of the cockroach Blattella germanica.
 #cross-references MIM:93238692
 #accession S30338
 #molecule_type mRNA
 #residues 1-856 #label MAR
 CLASSIFICATION #superfamily hydroxymethylglutaryl-CoA reductase (NADPH)
 KEYWORDS NADP; oxidoreductase
 SUMMARY #length 856 #molecular-weight 93106 #checksum 6833
 Query Match 1.94; Score 178; DB 2; Length 856;
 Best Local Similarity 19.9%; Pred. No. 6.49e-09;
 Matches 29; Conservative 63; Mismatches 51; Indels 3; Gaps 3;

Db 73 YSYOFCLOKRLQSKYIIGAGLFTVSSFFVSSVINFGLSDVS-DKDALPFFLLID 131
 QY 319 YVFSVRKRIDVRSRFLALCSVITTAGSLAMSGLCFEFGITLSLQSKDIFPVLIVLG 378
 Db 132 LSKATVLAQFALSSQSD-EVKHNIRAGIAMGPIITLDIVETLVIVGMISGRVLEV 190
 QY 379 LENSIVITK-SVVSMDTFDKIRVAQALSKGWHISKTLLEITLITGLATFVPVIOE 437
 Db 191 LCCFACMSVIVVVFMTFPACLSL 216
 QY 438 FCFIYVGLSDFMQLFSTILAM 463
 RESULT 9
 ENTRY A35728 #type complete
 TITLE hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34) - African clawed frog
 ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
 DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 ACCESSIONS A35728
 REFERENCE A35728
 #authors Chen, H.; Shapiro, D.J.
 #journal J. Biol. Chem. (1990) 265:4622-4629
 #title Nucleotide sequence and estrogen induction of Xenopus laevis 3-hydroxy-3-methylglutaryl-coenzyme A reductase.
 #cross-references MIM:90170974
 #accession A35728
 #status preliminary
 #molecule_type mRNA
 #residues 1-883 #label CHE
 CLASSIFICATION #superfamily hydroxymethylglutaryl-CoA reductase (NADPH)
 KEYWORDS NADP; oxidoreductase; transmembrane protein
 SUMMARY #length 883 #molecular-weight 96719 #checksum 6785
 Query Match 1.9%; Score 179; DB 1; Length 883;
 Best Local Similarity 18.7%; Pred. No. 4.64e-09;
 Matches 31; Conservative 68; Mismatches 62; Indels 5; Gaps 5;
 Db 61 DIILITRICALIYTFQFONLRQSGKIIGAGLFTTSSFFVS-TVVIHF-LDKEL 118
 QY 305 ELVPTVAVMLFAVYVFSRKRIDVRSRFLALCSVITTAGSLAMSGLCFEFGITLSL 364
 Db 119 TGLNEALPFFLLIDLSRASLAKFALSSNSODEVRENIRAGMAILGPTTEALVECV 178
 QY 365 QS-KDIFPVLIVLGLNSLVITKSVSMDETFDKIRVAQALSKGWHISKTLLEITI 423
 Db 179 IGVGTMSGVROLEIMCCFGCMSVLANYFAFMTFFPACVSLVELSR 224
 QY 424 LITGLATFVPVIOEFCIFAIYGLSDFMQLFST-I-LAMNIKR 467
 RESULT 10
 ENTRY S32572 #type complete
 TITLE hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34) - fruit fly (Drosophila melanogaster)
 ORGANISM #formal_name Drosophila melanogaster
 DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 11-Jun-1999
 ACCESSIONS S32572
 REFERENCE S32572
 #authors Gertler, F.B.; Chiu, C.Y.; Richter-Mann, L.; Chin, D.J.
 #journal Mol. Cell. Biol. (1988) 8:2713-2721
 #title Developmental and metabolic regulation of the Drosophila melanogaster 3-hydroxy-3-methylglutaryl coenzyme A reductase.
 #cross-references MIM:88302188
 #accession S32572
 #status preliminary
 #molecule_type DNA
 #residues 1-916 #label GER


```

Db 332 SYCNFGKQFFLEAAV-MITDILLFSFVAITLTKLEMRNRNARD-DYRKVLI 383
OY 432 V--PV-IOEFCIFAVIGLSDPMLQMLFSTIIAMNIRKREYTAFAAKHLPKML 482

RESULT 12
ENTRY B71215 #type complete
TITLE hypothetical protein PH1966 - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 26-Aug-1999

ACCESSIONS B71215
REFERENCE A71000
AUTHORS Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.;
Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
Hosokawa, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
Nakazawa, H.; Takamiya, M.; Ofuku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Roub, F.T.;
Horikoshi, K.; Masuch, Y.; Shizuya, H.; Kikuchi, H.
#journal #title Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
OT3.
#cross-references MIMD:98344137
#accession B71215
#status preliminary: nucleic acid sequence not shown;
translation not shown

#molecule_type DNA
##residues 1-293 ##label KAW
##cross-references GB:AP000007; NID:g3256134; PID:d1032056; PID:g325843
##experimental_source strain OT3
##note this accession replaces an interim accession for a
sequence replaced by GenBank

GENETICS PH1986
CLASSIFICATION #superfamily Methanobacterium thermoautotrophicum
protein-export membrane protein secf
SUMMARY #length 293 #molecular_weight 32209 #checksum 3965

Query Match 1.7%; Score 154; DB 2; Length 293;
Best local Similarity 25.5%; Pred. No.1.58e-05;
Matches 40; Conservative 52; Mismatches 55; Inde1s 10; Gaps 10;

Db 142 IVAFICMAVVEFLFRRV-VPSLIVFSAFSDMTVAVALMDISIEUS-QAT-IALIIM 198
OY 315 LVFAYVYFVRKIDVFRSRLALACVITTAGSIAMSLGICFFGLTISLQSKDIFPLYV 374
Db 199 LIGSVDSN-LLTTRLLRKE-FSVEAYVSSL-RGTFMSTTIGALASIMI-FST-AK 254
OY 375 ILV-GLSNSIVITKVSXMDFTDVIRVAQALSKEGWHISKTLTETITITIGIATVP 433
Db 255 VIDIASVLTEGLADFM-NTWIFNGAVRLIKRRE 290
OY 434 VIOEFCIFAVIGLSDPMLQMLFST-IIAMNIRKE 469

RESULT 13
ENTRY T16607 #type complete
TITLE hypothetical protein K1082.1 - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

ACCESSIONS T16607
REFERENCE T16545
AUTHORS Miller, N.
#submission submitted to the EMBL Data Library, June 1995
#description The sequence of C. elegans cosmid K1082.
#accession T16607
#status preliminary: translated from GB/EMBL/DBJ
#molecule_type DNA
##residues 1-701 ##label MIM

```

```
#experimental_source strain Bristol N2
```

PIDN:AAA68258.1; CESP:k10B2.1

```
GENETICS
#gene      CESP:k10B2.1
#ntons     78/3; 125/1; 183/2; 281/3; 404/3; 551/3; 668/3
SUMMARY    #length 701 #molecular_weight 80319 #checksum 5749
```

Query Match 1.7%; Score 155; DB 2; Length 701;
Best Local Similarity 26.0%; Pred.No.1.16e-05;

Matches 33; Conservative 40; Mismatches 48; Indels 6; Gaps 6.

Dc 371 ITIRRLVGHRAAVVVDFEDDRYIYASAGDRTIKWSMDLEFVRTLAGHRGIACLOYR 430
|||:::||:: ||:: ||:: |::||::||:: ||:: ||:: ||:
Oy 1005 IITLBEVRALHQPIRCMGQVNVDMPFTSGDTILTKYYCLNKNSDVEXTLLRHGCGPTCL-FV 1063

Dd 431 GR-LV-VSGS -SDNTIR-LMDIHSCVCLEARLEGHEELTCIFREDKRIYSGAHYDKRKIVW 486
Oy 1064 DRWGPGTGGSGSDGLLCWMDLFTGCACYNIQAHDGVAWSCAACAPSYVISLGTERICIW 1123

Dc 487 D-QAAIL 492
Oy 1124 ERFGNL 1130

```
RESULT   14
ENTRY    S62507          #type fragment
TITLE    hypothetical protein i (cosmid c29E6) - fission yeast
ORGANISM Schizosaccharomyces pombe) (fragment)
DATE     12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change
                24-Sep-1998 .

ACCESSIONS
REFERENCE S62507
           Jones, L.; Murphy, L.; McNeill, A.; Simpson, I.; Harris, D.
#authors submitted to the EMBL Data Library, October 1995
#submission S62507
#accession preliminary
         ##status
        ##molecule_type DNA
        ##residues       1-267 ##label JON
GENETICS ##cross-references EMBL:Z66525; NID:g1044926; PID:g1044927 *
CLASSIFICATION #superfamily WD repeat homology
FEATURE        #domain WD repeat homology #label WD1\
               16-51             #domain WD repeat homology #label WD2
              104-137            #length 267 #checksum 7169
SUMMARY        length 267 #checkum 7169
```

Query Match 1.6%; Score 145; DB 2; Length 267;
Best Local Similarity 26.8%; Pred.No.2.55e-04;

Matches 33; Conservative 35; Mismatches 52; Indels 3; Gaps 3.

Dc 61 FGHTDNVLGVVSENYIISSRDHARWRWDATSPNAACHYLGRHLASNVSVOYSKT 120
|||:::||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:
Oy 1012 LAHQPLITCMOVNDNMVFTSODRTLKYCIN-RSDVEYTTLHGCGPYTCL-FVD-RWPQ 1068

Dd 121 GLTVASSDRLFRLTDITTHGCIIRITHAHORIGACAQNGKFVIYSGSDLTTRIFEASSG 180
|||:::||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:
Oy 1069 GTGGSGSDGLLCWMLDTGTGACMYNQIAHDGAVSLCAAPSIVIISLGDERTCYWERFFQG 1128

Dc 181 KLL 183
Oy 1129 NL 1131

```
RESULT   15
ENTRY    E75029          #type complete
TITLE    Protein-export membrane protein secf related PAB1176 -
ORGANISM Pyrococcus abyssi (strain Orsay)
DATE     20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
                20-Aug-1999
ACCESSIONS E75029
```

```

REFERENCE      A75001
#authors       anonymous, Genoscope
#submission    submitted to the EMBL Data Library, July 1999
#description    Pyrococcus abyssi genome sequence; Insights Into archaeal
                Chromosome structure and evolution.
#accession     E75029
#status        Preliminary
#molecule_type DNA
##residues     1-308 ##label KAM
##cross-references GB:A748288; GB:AL096836; NID:g458960.
                PIDN:CABS0675.1; PID:e1516574; FID:g5459189
#experimental_source strain Orsay

GENETICS
#gene          PAB1176
SUMMARY        #length 308 #molecular_weight 33806 #checksum 9284

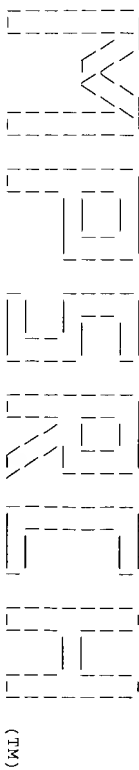
Query Match      1.6%; Score 150; DB 2; Length 308;
Best Local Similarity 23.8%; Pred. No. 5,50e-05;
Matches 40; Conservative 51; Mismatches 53; Indels 11; Gaps 11;

Db 157 ITVAFIGMA-IWVLEFLPIVPSTLVFSAFSDMVAVMLNIFGIELS-QAT-ITALLM 213
:: : :: | | :|: | : | : : : : : | : | : | : : | :
Oy 317 FAIVYFSVRKIDVRS-RFL-LALCSVTITNGSLAMSGLCFEFGLLISQSKDIPFYLV 374
:: : :: | | :|: | : | : : : : : | : | : | : : | :
Db 214 LIGSVDSNNILFTTRLKRKE-FTVEBARYSSL-RTGFTNSTTTGALISLWL-FST-AK 269
:: : :: | | :|: | : | : : : : : | : | : | : : | :
Oy 375 ILV-GLNSLTIVTSVSMDETPEVKIRVAQALSKEGMHSKTLTEIITLTIGLATFPV 433
:: : :: | | :|: | : | : : : : : | : | : | : : | :

Db 270 VIDIASVLVGLADFPNTMIINAGVLRMYIAIKR 304
| : : : : | : | : | : | : | : | : | : | : | : | :
Oy 434 VIQEFICAIYGLSDERLOMRLSTLIANNI-KR 467

```

```
Search completed: Sun Sep 3 12:45:00 2000
Job time : 106 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MATCH-PP protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:39:57 2000; Maspar time 87.99 Seconds
974.755 Million cell updates/sec

Tabular output not generated.

Title: >US-09-332-522B-6
Description: (1-1237) from US09332522B.Pep
Perfect Score: 9226
Sequence: 1 MKNKTKSKHFKAMKFWPAC.....CDYGNELRVRFPIVADKCH 1237

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 59334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb12
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-phc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 55.496; Variance 103.793; scale 0.535

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	1265	13.7	1277	4	KIAA0199 PROTEIN (FRAG	2.05e-238
2	1248	13.5	1276	11	SREBP CLEAVAGE ACTIVAT	1.23e-234
3	665	7.2	1032	5	D2013.8 PROTEIN.	4.43e-107
4	297	3.2	1086	3	HYPOHETICAL 125.1 KD	1.29e-31
5	271	2.9	1170	3	HYPOHETICAL 132.6 KD	1.04e-26
6	248	2.7	608	11	PATCHED (FRAGMENT).	1.84e-22
7	245	2.6	800	5	Y38F1A.3 PROTEIN.	4.25e-22
8	244	2.6	587	5	SEL-10.	9.82e-22
9	237	2.6	1278	4	NIEMANN-PICK C DISEASE	1.81e-20
10	230	2.5	881	5	T21H3.2 PROTEIN.	3.27e-19
11	229	2.5	1146	4	PATCHED 2.	4.83e-19
12	229	2.5	1203	4	RECEPTOR PROTEIN PATCH	4.93e-19
13	229	2.5	1203	4	RECEPTOR PROTEIN PATCH	4.93e-19
14	234	2.5	1243	13	PATCHED-2 PROTEIN.	6.28e-20
15	224	2.4	1278	11	NPCL.	3.82e-20
16	218	2.4	1318	5	PUTATIVE HEDGHOG RECE	4.37e-17
17	204	2.2	840	5	T07H8.6 PROTEIN.	1.19e-14
18	205	2.2	933	5	SIMILAR TO DROSOPHILA	7.99e-15
19	196	2.1	889	5	SIMILAR TO DROSOPHILA	2.76e-13
20	191	2.1	936	5	Y18D10A.7 PROTEIN.	1.92e-12

21	172	1.6	820	5	044978	F56C11.2 PROTEIN.	2.57e-09
22	171	1.6	803	5	076819	3-HYDROXY-3-METHYLGLUT	3.72e-09
23	174	1.6	1003	5	09XMR9	Y39A1B.2 PROTEIN.	1.22e-09
24	177	1.6	1053	3	074425	3-HYDROXY-3-METHYLGLUT	3.98e-10
25	164	1.6	845	5	P91184	SIMILAR TO C. ELIGANS	4.82e-08
26	154	1.6	293	1	057717	293KA LONG HYPOHETICA	1.72e-06
27	156	1.7	510	5	044382	SLIMB.	8.48e-07
28	160	1.7	550	4	094366	KIAA0696 PROTEIN (FRAG	2.04e-07
29	158	1.7	877	5	021883	RO9H10.4 PROTEIN.	4.17e-07
30	145	1.6	386	5	09XWU3	Y41CA4.11 PROTEIN.	3.88e-05
31	149	1.6	569	4	09Y213	BETA-TRANSDUCIN REPEAT	9.83e-06
32	149	1.6	569	11	092159	BETA-TRANSDUCIN REPEAT	9.83e-06
33	149	1.6	605	4	09Y297	B-TCP VARIANT E3RS-1K	9.83e-06
34	145	1.6	1015	5	009938	HYPOHETICAL 116.3 KD	3.88e-05
35	146	1.6	1405	5	009614	HYPOHETICAL 155.4 KD	2.76e-05
36	142	1.5	155	8	079069	NADH DEHYDROGENASE 1 (1.07e-04
37	134	1.5	376	13	091652	GENE 16.	1.52e-03
38	136	1.5	577	5	097334	PFC0100C PROTEIN.	7.89e-04
39	134	1.5	690	5	061585	KATININ P80 SUBUNIT.	1.52e-03
40	140	1.5	750	1	029039	CONSERVED HYPOHETICAL	2.10e-04
41	134	1.5	888	2	083769	ANTIBIOTIC TRANSPORT P	1.52e-04
42	139	1.5	956	5	018129	SIMILAR TO DROSOPHILA	2.93e-04
43	132	1.4	155	8	079072	NADH DEHYDROGENASE 1 (2.90e-03
44	133	1.4	714	5	009540	HYPOHETICAL 79.1 KD P	2.10e-03
45	131	1.4	955	5	019945	F31F6.5 PROTEIN.	4.00e-03

ALIGNMENTS

RESULT 1
ID 012770 PRELIMINARY; PRT: 1277 AA.

AC 012770;
DT 01-NOV-1996 (Tribble, 01, Created)
DT 01-NOV-1996 (Tribble, 01, Last sequence update)
DT 01-MAY-1999 (Tribble, 10, Last annotation update)
DE KIAA0199 PROTEIN (FRAGMENT).
GN KIAA0199.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96281124.
RA MNGASE.T.; SEKI N.; ISHIKAWA K.; TANAKA A.; NOMURA N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1."
RL DNA Res. 3:17-24(1996).
DR EMBL: D83782; BAA12111.1; .
DR PFM: PFM0400; WD40; 3.
FT NON_TER
SQ SEQUENCE 1277 AA: 139495 MW; 10EB5099 CRC32;

Query Match 13.7%; Score 1265; DB 4; Length 1277;

Best local Similarity 34.5%; Pred. No. 2.05e-238;
Matches 277; Conservative 194; Mismatches 262; Indels 71; Gaps 44;

Db	64	PPVDSRKQGEPTQPEWYGAFA-VYQIEFKSVFPMKMLAVDFRSPISRAQ	122
Qy	81	POVINSSTTRSPPLPMAQSSPAFFVQIITRTSVLPMTGQMLDARRAPLHEVFK	140
Db	123	LVEIRNRVLRDSSGIRSELELCQVYDLPGLR-KLRNLRLPHGCLLSGNQWQNW	181
Qy	141	LLEIVRNH--OSSENKRLIEHVDVKKGTGQDQIDPEYGCCLLSPANLMTQNSQ	198
Db	182	RFHADPDICGTHHEPKTQTS-ATLKDLFGVKGKSGSLVT-R-K-RMVSYTITLV	237
Qy	199	NFTSDTINLNTIIFQY-N-LQSKVSAEMFGPLMDGTGKRRLPARSIIQYALTLF	256
Db	238	FOHYHAKFLSLRLRLMLAPS-PNCSIRAE---SLVHVFKEEIGVAELIPVTYIIL	293
Qy	257	LKHDMDEYLDLTKELRLRHPPLPLASASAEPTIYIFYPGEYRMWELPVYVAFMLV	316

[illegible]

Query Match	13.5%;	Score 1248;	DB 11;	Length 1276;
Best Local Similarity	34.3%;	Pred. No. 1.23e-234;		
Matches	280;	Conservative 187;	Mismatches 287;	Indels 63; Gaps 43

```

Db      47 LPLPETGPFVEFSTPVKXDYSP-PPV-DSOHKQGESEQEPENIVGAPYA-YIQIIFVKSYS 103
      |::||::||::||::||::||::||::||::||::||::||::||::||::||
QY     60 LNTSSTTPQEHHPGSGEPWPPEQVLNSTTDRSPPLLPMAQSSPAFFYQQITLRSVL 119

```

Dd	104	PMHNHLAVDVEFRLPSRAPOLVEIRNNHVLRODSSGKSLSEEOVLQYTDLLPGHR-KLRN	162
Qy	120	PMTEGMQMDAFRAPHLEHVKEVLEIVANH--QSEKNRTLEHNCHLDVNVKRGHGLDQ	177
Dd	163	LLPEHGCLLSPGNFWMQDMEFRHADPDIOGTIHQHEPKTLQTS-ATLKDLFGVPGRKYS	221
Qy	178	IFPEYGCCLLSPALMTQNSQNFRTDNIINLTIFYQH-N-LOKSKVAAEMLEGLPMQDT	235
Dd	222	GVSLYT-R-K-RVTSYITTLVFOQYHAKFSLSLRARLMLHPs-PNOSLAEN---LVHV	274
Qy	236	GFKRPLPLARSRITIOVALTFLEKNDMEYDITLKEKLLRHPRPLUPLASASAEETTTYI	295
Dd	275	HFKEBIGIAELIPVTYIIILFAIYFSTRKIDWKSCKGATLAAVTVSSILMSVGLC	334
Qy	296	FYPGEYRMELVPYTVAPMLFAVYVSXKIDYFRFRFLALCSVTTTNGSLAMSGLC	355
Dd	335	TLFGTLPTLNGEIFPVLVVYIGLENNVLTKSVYSTPVDELVKRLIAQGLSSESWSIMK	394
Qy	356	FFFGTITLSQKDDIFPLVILVIGLENSIVTKSVYSDMEFDVYKIRVAOALKEGWHSIK	415
Dd	395	NVAHELGILIGITTLVPAIOECLFVAVGCLVSDFLOMFETTVLSIDIRMLADLNK	455
Qy	416	TLTEITLITLTGLATFVYDIOEFCIFAIVGLISDFMLOMLEFTILAMNKRTEYTAEK	475
Dd	455	RLPPE--SCLEPSA-----K-----P-VGR--PAYER-Q-----LA-VRPA-MPTHTL-Q	491
Qy	476	HLPKMLUSTCTGAGRODFRFGAARLPBPYPVGTGFQNSQHPKICFDPDASVSRTSVN	533
Dd	492	PSSEFNLRLPKRLVYIFFLARTLRAORLINAQTVVWIGILVYTDPAGLRTYLAOVTEQS	551
Qy	536	GHSSEQGRIPRIKIVNEMARTREFQRAFMIMTWICSHY-N-SG---YLEQDFMSQ	590
Dd	552	PLGEGSLCPMVPBGVLPAKRPD-PAR-STPPPAKPLPENQIYPG-ELDEHAPADGVH	608
Qy	591	N-GTMT-ATLELQIR-RLQAGAGVASFEGEQADQGRATSAPSGSGSTPIKAPLADIN	647
Dd	609	DSRAPEYMWGEDE-ELMRRLSFHFHWPLTFNXYITLAKRISLPIVPIRLINPOEAL	665
Qy	648	ET--AEEMRLRIPSEDLVYFLSNFHWSTIMKOIYISIGHVITLPLTRISHAIAPLAT	706
Dd	722	L--V-L-LILLCYRYLCPBNHQPGGAGRRRRGELPCDDIGYAPPETEIVPLVLRGLMD	778
Qy	766	LCGISIFVLCYTMVYFRCLITRNYAEMRSSMHSE--APYKOTEOILEGVPQTQIAGHKHR	824
Dd	779	TECLASQGMILLVSCCLAGQVCMVAQOAGDCLLRIPR	815
Qy	825	TECLVSDGAYITISCCLGQIRVWPAQSGEOLTSISR	861

RESULT	3	PRELIMINARY;	PRT; 1032 AA.
ID	Q18968		
AC	Q18968;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1998 (TREMBLrel. 07, Last sequence update)		
DT	01-MAY-1999 (TREMBLrel. 10, Last annotation update)		
DE	D2013.8 PROTEIN.		
GN	D2013.8.		
OS	<i>Caenorhabditis elegans</i> .		
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabdilitida;		
OC	Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabditis		
NC	[1]		
RP	SEQUENCE FROM N.A.		
RA	MORTIMORE B.;		
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.		

RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,

DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)
DE PATCHED 2.
GN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP
RP SEQUENCE FROM N.A.
RA MEDLINE; 99135908.
RA SMYTH I., NARANG M.A., EVANS T., HEIMANN C., NAKAMURA Y.,
RA CHENEVIX-TRECH G., PIETSCH T., WICKING C., MAINWRIGHT B.J.,
RT "Isolation and characterization of human patched 2 (PCH2), a putative
RT tumour suppressor gene inbasal cell carcinoma and medulloblastoma on
RT chromosome 1p32." 6:291-297(1999).
RL Hum. Mol.Genet. 8:291-297(1999).
RL EMBL; AF087651; AAD23553.1;
SQ
SQ SEQUENCE 1203 AA; 130498 MW; 226B2BAB CRC32;

Query Match	2.58;	Score 229;	DB 4;	Length 1203;
Best Local Similarity	22.28;	Pred. No. 4.93e-19;		
Matches	35;	Conservative	66;	Mismatches 53;
			Indels 4;	Gaps 4;

Ddb 401 GYLLMLAYACTMTLMRNDCAOSGSGVGLAVLVALAVASGLICALLGITENNAATQVLP 460
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 312 AFMLVFAYVYSVRKIDVFRSEFLALCSVITTAGSLANSIGLCFFFGLTISLSQSDIFP 371

Db 461 FLALGIGDVDFLLAHAFTEALPGT-PLQERMGCLORTGTSVLTSSINNAFLMAALY 519
::
Qy 372 YLVAVLGLSVITKSV-SNDETFDKIRVAQALSKEGMHISKTLTTETT-ILTIGLA 429

Dd 520 P-IPALRAFLSLQAAIVVCGTFVAVMLEVPAILSDLR 556
 :|::||:||::||:
Dy 430 TFPVTIQEFCIFAIVGLISDFMLQMLLFTLILAMNIKR 467

Figure S9. Sequence alignment between Dd and Dy proteins.

RESULT 13
ID O95341
PRELIMINARY;
PRT; 1203 AA

DT	01-MAY-1999	(TEMBLrel, 10, Created)
DT	01-MAY-1999	(TEMBLrel, 10, Last sequence update)
DT	01-MAY-1999	(TEMBLrel, 10, Last annotation update)

22. INDELION ENOILIN ENCHID 2.
GN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
PTCH2.

CC	ECURINERIA; PRIMATES; CATALANINI;	ROMULINAE; ROMULI
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE; 99030620.	

RT protein family.";
RA "Characterization of two patched receptors for the vertebrate hedgehog
ROSENTHAL A., DE SAUVAGE F.J.;
RA CARPENTER D., STONE D.M., BRUSH C., RIAN A., ANTONINI M., FRANZ G.,
RA ROSENTHAL A., DE SAUVAGE F.J.;
RT protein family.";

DR EMBL; AF091501; AAC79847.1; -.
 KM Receptor.
 SO SEQUENCE 1203 AA: 130573 MW: 21829476a CRC32:
 RL PROC NAtl. Acad. Sci. U.S.A. 95:13630-13634 (1998)

Query Match	2.58;	Score 229;	DB 4;	Length 1203;
Best Local Similarity	22.28;	Pred. No. 4.93e-19;		
Matches	35;	Conservative	66;	Mismatches 53;
				Indels 4;
				Gaps 4;

Dd 401 GYLEMLAYACVETMLRNDCAOSGSGVLGLVALAVASGLGCLALGITFNAAITQYLP 460
:::| | : : : : : : : : : : : : : : :
Qy 312 AFMLVFAYVVYSVRKIDVFRSRFLALCSVITTTNGSLAMSLGLCPFFGLTISLQSKDIFP 371

D_b 461 FLAIGCVDDVFLLAHAFTEALPGT-PLÖERMGECLQRPSTSVITISINNAFFLMAALY 519
 :: :: :: :: :: : | :: :: : | :: :: : |
Q_y 372 YLVAVLGLENSLVITFSVV-SNDETFDVKIRVAQALSKEGWHISKTLTTETT-LIVIGLA 429

Db 520 P-IPALRAEFLQAAIVVGCTFEAVAMVPEPAILSLDLRR 566
 :|::||::||: | ||||:||||:||||:

..

QY 430 TFVPVIOEFCIFAIVGLSDFMLQMLLESTILAMNIKR 467

RESULT	14
ID	Q9W6T6
	PRELIMINARY;
	PRT; 1243 AA

DT	01-NOV-1999	(TREMBlurel, 12, Created)
DT	01-NOV-1999	(TREMBlurel, 12, Last sequence update)
DT	01-NOV-1999	(TREMBlurel, 12, Last annotation update)

GN PTCL.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;

OC Cyprinoidae; Cyprinidae; Rasbora; Danio.
RN [1]
RP SEQUENCE FROM N.A.

RT "Characterisation of the second ptc gene in zebrafish."
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases
 EMBL: AJ007742; CAB39726.1; -
 DR

Query Match	2.58;	Score 234;	DB 13;	Length 1243;
Best Local Similarity	22.28;	Pred. No. 6.28e-20;		

	MALICES	CONSERVATIVE	07,	MISMALICES	07,	INOCES	4,	SAPS	2
Db	428	VIRIAGYLIMLAACTLMRWDCAKSQGAVGLIVLTLSVAAGLGICSLGISFNMA	487						
:	:	: : : : :	:	:	:	:	:	:	:

Db 488 TTQVLPFLALGVGDVDFLLAHAFSENGQNRIPEEDRTGCELRGTASVVLTSISNVTA 547

QY 300 SKDIFPLVLVLGNSLVIAH-SVWSMDEIFDVAKRVAQALSAESGWHIS

423 ILLIGATEPVIOEFCIPIAVGULSDPMLQMLLESIILAMNINKI

ID	035604	PRELIMINARY;	PRT;	12/8 AA.
AC	035604;			
DT	01-JAN-1998	(TREMblrel. 05, Created)		
DT	01-JAN-1998	(TREMblrel. 05, last sequence update)		

Accession	Species	Accession	Species
BT	01-NOV-1999 (ENSEMBLrel. 12, last annotation update)	BT	01-NOV-1999 (ENSEMBLrel. 12, last annotation update)
DE	NPCL.	DE	NPCL.
GN	NPCL.	GN	NPCL.
OS	Mus musculus (Mouse).	OS	Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 PR SEQUENCE FROM N A

RX MEDLINE; 97362324.
RA LOFUS S.K., MORRIS J.A., CARSTEA E.D., GU J.Z., CUMMINGS
RA BROWN A., ELLISON J., OHNO K., ROSENFELD M.A., TAGLE D.A.,
RA DRENGELUND D.C., DAVAN E.T.

RT "Murine model of Niemann-Pick C disease: mutation in a cholesterol
RT homeostasis gene."; Science 277:232-235(1997).
RL PubMed 9166630

DR MGD; MGI:1097712; Npc1.
SQ SEQUENCE 1278 AA; 142887 MW; 072DA88B CRC32;

Best Local Similarity	26.3%;	Pred. No. 3.82e-18;
Matches	45;	Conservative 61; Mismatches 54; Indels 11; Gaps 9;

306 LVPYTVAFMLV-FAIYVF-SVRKIDYRSRFLALCSVITTAGSLAMSLGCEFFFGILTIS 363

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

OY | ::|::|: |::| ::: : | :: : | :: | ::|: |:
364 LQSKDIEPYIWLIVGLNSLIVITKSVSMDETFDKIRVAQALSKE-GWHISKTL-LT-- 419
DB 742 SETSAFFEGALUSMPAVHTFSLFAGMAVLIDFLLQITCFVSLIGDIKROE 792
| : : : | : : | : : | : : | : : | : : | :
OY 420 -ETITLITGLATFVPVIOEFCIFAIVGLSDPMLQMLFSTILAMNKRTE 469

Search completed: Sun Sep 3 12:42:57 2000
Job time : 180 secs.

THIS PAGE BLANK (USPTO)

NWSENF
(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpstrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:38:30 2000; Maspar time 37.66 Seconds

Tabular output not generated. 1017.918 Million cell updates/sec

Title: >US-09-332-522B-6
Description: (1-1237) from US09332522B.pep
Perfect Score: 9226
Sequence: 1 MNKTKSKHFKAWKFWPAC.....CDYGNIRVVRPIVADKCH 1237

Scoring table:
PAM 150
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 56.799; Variance 99.679; scale 0.570

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	246	2.7	1220	1	PTC1_BRARE	1.08e-23
2	244	2.6	579	1	SEL-10 PROTEIN.	2.61e-23
3	240	2.6	1434	1	PTC1_MOUSE	1.52e-22
4	239	2.6	1442	1	PTC1_CHICK	2.37e-22
5	240	2.6	1447	1	PTC1_HUMAN	1.52e-22
6	223	2.4	1182	1	PATCHED PROTEIN HOMO	2.51e-19
7	212	2.3	932	1	3-HYDROXY-3-METHYLGLUT	2.75e-17
8	182	2.0	703	1	WD-REPEAT PROTEIN POP2	6.46e-12
9	180	2.0	732	1	KHMB_DICD1	1.43e-11
10	183	2.0	887	1	3-HYDROXY-3-METHYLGLUT	4.33e-12
11	183	2.0	887	1	3-HYDROXY-3-METHYLGLUT	4.33e-12
12	183	2.0	887	1	3-HYDROXY-3-METHYLGLUT	4.33e-12
13	183	2.0	888	1	3-HYDROXY-3-METHYLGLUT	4.33e-12
14	183	2.0	888	1	3-HYDROXY-3-METHYLGLUT	4.33e-12
15	178	1.9	856	1	3-HYDROXY-3-METHYLGLUT	2.13e-11
16	179	1.9	883	1	3-HYDROXY-3-METHYLGLUT	2.13e-11
17	177	1.9	916	1	3-HYDROXY-3-METHYLGLUT	4.71e-11
18	177	1.9	1053	1	3-HYDROXY-3-METHYLGLUT	4.71e-11
19	162	1.8	332	1	GLP1_DI	1.57e-08
20	162	1.8	605	1	GUANINE NUCLEOTIDE-BIN	1.57e-08
21	155	1.7	701	1	HYPOHETICAL 67.1 KDA	2.17e-07
22	145	1.6	267	1	HYPOHETICAL TRP-ASP R	8.31e-06
23	147	1.6	479	1	PP1/PP2A PHOSPHATASES	4.05e-06

ALIGNMENTS

RESULT	ID	1	STANDARD:	PRT:	1220 AA.
AC	Q98864;				
DT	15-FEB-2000 (Rel. 39, Created)				
DT	15-FEB-2000 (Rel. 39, Last sequence update)				
DT	15-FEB-2000 (Rel. 39, Last annotation update)				
DE	PATCHED PROTEIN HOMOLOG 1 (PATCHED 1) (PTC1).				
GN	PTC1.				
OS	Brachydanio rerio (zebrafish) (zebra dānio).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;				
OC	Cypriniformes; Cyprinidae; Rasbora; Danio.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-EMBRYO;				
RX	MEDLINE; 96379744.				
RA	Concordet J.-P., Lewis K.E., Moore J.W., Goodrich L.V., Johnson R.L.,				
RA	Scott M.P., Ingham P.W.;				
RT	"Spatial regulation of the zebrafish patched homologue reflects the				
RT	roles of sonic hedgehog and protein kinase A in neural tube and somite				
RT	patterning."				
RT	Development 122:2835-2846(1996).				
RL	-1- FUNCTION: ACTS AS A RECEPTOR FOR SONIC HEDGEHOG (SHH), INDIAN				
CC	HEDGEHOG (IHH) AND DESERT HEDGEHOG (DHH). ASSOCIATES WITH THE				
CC	SMOOTHENED PROTEIN (SMO) TO TRANSDUCE THE HEDGEHOG'S PROTEINS				
CC	SIGNAL (BY SIMILARITY).				
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.				
CC	-1- TISSUE SPECIFICITY: DETECTED IN EMBRYONIC PRESOMITIC MESODERM,				
CC	NEUROECODERM, TISSUE SURROUNDING THE NOTOCHORD, VENTRAL NEURAL				
CC	TUBE.				
CC	-1- DEVELOPMENTAL STAGE: AT ALL STAGES, EXPRESSION CORRESPONDS TO THE				
CC	LOCALIZATION OF SHH. FIRST DETECTED DURING GASTRULATION. BY 36				
CC	HOURS, PTC1 APPEARS IN THE FIRST BRANCHIAL ARCH AND THE POSTERIOR				
CC	MESONEURUM OF THE FIN BUD; BY 48 HOURS, IN THE HINDBRAIN AND				
CC	FOREBUD.				
CC	-1- INDUCTION: ACTIVATED BY SONIC HEDGEHOG.				
CC	-1- PTM: GLYCOSYLATION IS NECESSARY FOR SHH BINDING (BY SIMILARITY).				
CC	-1- SIMILARITY: BELONGS TO THE PATCHED FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				

DR	EMBL: X98883; CAAG7386.1; -	PTC1.
DR	ZFIN; ZDB-GENE-980526-44; -	
KW	Receptor, Transmembrane; Glycoprotein.	
FT	DOMAIN	1 84
FT	TRANSMEM	85 105
FT	DOMAIN	106 419
FT	TRANSMEM	420 440
FT	DOMAIN	441 449
FT	TRANSMEM	450 470
FT	DOMAIN	471 484
FT	TRANSMEM	485 505
FT	DOMAIN	506 528
FT	TRANSMEM	529 549
FT	DOMAIN	550 558
FT	TRANSMEM	559 579
FT	DOMAIN	580 739
FT	TRANSMEM	740 760
FT	DOMAIN	761 1016
FT	TRANSMEM	1017 1037
FT	DOMAIN	1038 1044
FT	TRANSMEM	1045 1065
FT	DOMAIN	1066 1072
FT	TRANSMEM	1073 1093
FT	DOMAIN	1094 1110
FT	TRANSMEM	1111 1131
FT	DOMAIN	1132 1143
FT	TRANSMEM	1144 1164
FT	DOMAIN	1165 1220
FT	CARBOHYD	397 865
FT	CARBOHYD	866 888
FT	SEQUENCE	1220 AA: 135544 NM: D10A9D04115F532D CRC64;

	Query Match	2.7%	Score 246;	DB 1;	length 1220;	
	Best Local Similarity	23.9%	Pred. No.	1.08e-23;		
Matches	36;	Conservative	61;	Mismatches	58;	Indels 2; Gaps 2
Db	428	GILLMLAYACTMRLRMDCAKSGQAGLAGVLLVALSVNAGLGCSLGLSEFNATTVLP	487			
Oy	312	AFMLVFAYVYSVKRIDYFRSRFLALCISVTTTAGSLAMSGLCFEFPBLTTSLOSKDIFP	371			
Dd	488	SALGIGVDMMFLGHSTETRSNIPFEKRTGDCLRTGTSVALTSSVMNAFFMALVP	547			
Oy	372	YLVIIVGLENSLVIRKSIVSVMDEFDVXIKVAQAALSKRGWMIKTLLLE-TITLTIGLAT	430			
Dd	548	-IPALRAFSLOAAVVYVNFPMALLIPIAIIISLDIHRE	585			
Oy	431	FVPVIOEFICFAIVGLSDPFMIOMLFESTILAMNIKRE	469			
RESULT	2					
ID	SE10-CAEEL	STANDARD;	PRT;	579 AA.		
AC	OQ93794;					
DT	15-JUL-1998 (Rel. 36, Created)					
DT	15-DEC-1998 (Rel. 37, Last sequence update)					
DT	15-DEC-1998 (Rel. 37, Last annotation update)					
DE	SEL-10 PROTEIN.					
GN	SEL-10 OR F55B12.3.					
OS	Caenorhabditis elegans.					
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;					
OC	Rhabdilitidae; Peloderinae; Caenorhabditis.					
[1]						
SEQUENCE FROM N.A.						
STRAIN-BRISTOL N2;						
Sims M.;						
Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.						
(2)						
REVSIONS.						
STRAIN-BRISTOL N2;						
Jones S.J.M.;						
Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.						
-I- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).						
CC						

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL; J79757; CAB02129.1; -.
DR WORMPEP; F55H12.3; CE16120.
DR PFAM; PF00646; F-box; 1.
DR PFAM; PF00400; WD40; 7.
DR PRINTS; PRO0320; GPROTINRPT.
DR PROSITE; PS00678; WD_REPEATS; 5.
KW Repeat; WD repeat.
KW Repeat; WD repeat.
FT REPEAT 245 274 WD1.
FT REPEAT 286 316 WD2.
FT REPEAT 328 356 WD3.
FT REPEAT 368 396 WD4.
FT REPEAT 408 438 WD5.
FT REPEAT 453 481 WD6.
FT REPEAT 493 522 WD7.
SQ SEQUENCE 579 AA; 64275 MW; 0647245AF5964663 CRC64;

Query Match 2.6%; Score 244; DB 1; Length 579;
Best Local Similarity 38.1%; Pred. No. 2,61e-23;
Matches 48; Conservative 31; Mismatches 41; Indels 6; Gaps 5;

```

```

Db 245 GHBRVITCMQIHDDVLTGSDGNTLKWCIDKGEVMTLGHGQWTSQ-IS--OCGR 301
Oy 1013 AHQ-OPITCMQVNDMMFTVSGSDHLLKCYCLNKSSVEYTLHGHCQPV-TCLFVDRMDPGT 1070
Db 302 YIVSGSDRTKAVSWTVSDGSLHHTLOGHTSTVRGAMAGSLIVTGSRDPTLRWDVESGR 361
Oy 1071 -GGSQSGGLLCVMDLFTGACMYNIQAHDGAVSCLACAPSVYISLGTDERICVERFQGN 1129
Db 362 HLAHLH 367
Oy 1130 LLETTIN 1135

RESULT 3
ID PRC1 MOUSE STANDARD; PRT; 1434 AA.
AC Q61115.
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PATCHED PROTEIN HOMOLOG 1 (PC1) (PTC).
GN PCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE; 96176226.
RA Goodrich L.V., Johnson R.L., Milenkovic L., McMahon J.A., Scott M.P.;
RT "Conservation of the hedgehog/patched signaling pathway from flies to
RL mice: induction of a mouse patched gene by Hedgehog.";
RL Genes Dev. 10:301-312(1996).
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR SONIC HEDGEHOG (SHH), INDIAN
CC HEDGEHOG (IHH) AND DESERT HEDGEHOG (DHH). ASSOCIATES WITH THE
CC SMOOTHENED PROTEIN (SMO) TO TRANSDUCE THE HEDGEHOG'S PROTEINS
CC SIGNAL. SEEMS TO HAVE A TUMOR SUPPRESSOR FUNCTION, AS INACTIVATION
CC OF THIS PROTEIN IS PROBABLY A NECESSARY, IF NOT SUFFICIENT STEP
CC FOR TUMORIGENESIS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT LOW LEVEL AT 7 DAYS POST COITUM
CC (DPC), HIGHER LEVEL AT 11 AND 15 DPC, AND PERSISTS AT MODERATE
CC LEVELS AT 17 DPC. IN THE DEVELOPING EMBRYO, PTC IS FIRST DETECTED
CC WITHIN THE VENTRAL NEURAL TUBE AND LATER IN THE SOMITES AND LIMB
CC BUDS. EXPRESSION IN THE LIMB BUDS IS RESTRICTED TO THE POSTERIOR
CC ECTODERM SURROUNDING THE ZONE OF POLARIZING ACTIVITY. IN THE

```



```

Query March 2.0%; Score 183; DB 1; Length 887;
Best Local Similarity 18.7%; Pred Num 4,33e-12;
Matches 31; Conservative 68; Mismatches 62; Indels 5; Gaps 5;

Db      61 DIITITRCIAIITYPOFONROLGSKYILGIAGFTTSSFEVS-TVVIHF-LDKEL 118
       ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy      305 ELVPVTAFMVFAYFAIVSVSRKIDVFRSRLFALCVITTTGLSLMSGLCFEGLISTL 364
       ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db      119 TGLNALPFFELLIDLSASALFKAFALSNSODEVRNINARMAITGGTFPTLDALVELLV 178
       ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy      365 QS-KIIFFYLVLNGLENSLYTKSVSNDETFDVKIRYAQLASSEGWHISKTLLTEITI 423
       ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db      179 IGVGTMSGVROLETIMCCFCGMSVLANTFEVMTFPACYSLVLELRS 224
       ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy      424 LTIGATFVPVIGQECIFAIYAGLLSDFMLOMLLFST-I-LAMNIKR 467
       ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 11
AC      HMDH-CRIGR STANDARD: PRT: 887 AA.
AD      P00347;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DT      01-NOV-1997 (Rel. 35, last annotation update)
DE      3-HYDROXY-3-METHYLGUTARYL-COENZYME A REDUCTASE [EC 1.1.1.34] (HMG-COA
DE      REDUCTASE).
GN      HMCCR.
OS      Cricetus griseus (Chinese hamster).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC      Cricetulus.
RN      [1]
RX      SEQUENCE FROM N.A.
RX      MEDLINE: 84168178.
RA      Chin D.-J., Gall G., Russell D.W., Liscum L., Luskey K.L., Basu S.K.,
RA      Okayama H., Berg P., Goldstein J.L., Brown M.S.;
RA      "Nucleotide sequence of 3-hydroxy-3-methyl-glutaryl coenzyme A
RT      reductase, a glycoprotein of endoplasmic reticulum.";
RL      Nature 308:613-617(1984).
RN      [2]
RX      GENE STRUCTURE.
RX      MEDLINE: 84282111.
RA      Reynolds G.A., Basu S.K., Osborne T.F., Chin D.-J., Gall G.,
RA      Brown M.S., Goldstein J.L., Luskey K.L.;
RA      "HMG CoA reductase: a negatively regulated gene with unusual promoter
RT      and 5' untranslated regions.";
RL      Cell 38:275-285(1984).
RN      [3]
RX      DOMAINS.
RX      MEDLINE: 85080129.
RA      Liscum L., Finer-Moore J., Stroud R.M., Luskey K.L., Brown M.S.,
RA      Goldstein J.L.;
RA      "Domain structure of 3-hydroxy-3-methylglutaryl coenzyme A reductase,
RT      a glycoprotein of the endoplasmic reticulum.";
RL      J. Biol. Chem. 260:522-530(1985).
CC      -1- FUNCTION: THIS TRANSMEMBRANE GLYCOPROTEIN IS INVOLVED IN THE
CC      CONTROL OF CHOLESTEROL BIOSYNTHESIS. IT IS THE RATE-LIMITING
CC      ENZYME OF STEROID BIOSYNTHESIS
CC      -1- CATALYTIC ACTIVITY: (R)-MEVALONATE + COA + 2 NAD(P)+ = (S)-3-
CC      HYDROXY-3-METHYLGUTARYL-COA + 2 NADPH.
CC      -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
CC      -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
CC      RETICULUM.
CC      -1- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/)
CC      or send an email to license@isb-sib.ch).
CC      -----

```

DR	EMBL	L00183	AAA36989.1	-	JOINED.	
DR	EMBL	L00166	AAA36989.1	-	JOINED.	
DR	EMBL	L00169	AAA36989.1	-	JOINED.	
DR	EMBL	L00170	AAA36989.1	-	JOINED.	
DR	EMBL	L00171	AAA36989.1	-	JOINED.	
DR	EMBL	L00173	AAA36989.1	-	JOINED.	
DR	EMBL	L00176	AAA36989.1	-	JOINED.	
DR	EMBL	L00177	AAA36989.1	-	JOINED.	
DR	EMBL	L00178	AAA36989.1	-	JOINED.	
DR	EMBL	L00179	AAA36989.1	-	JOINED.	
DR	EMBL	L00180	AAA36989.1	-	JOINED.	
DR	EMBL	L00181	AAA36989.1	-	JOINED.	
DR	EMBL	L00182	AAA36989.1	-	JOINED.	
DR	EMBL	X00494	CAA25189.1	-	JOINED.	
DR	PIR	A00357	RDHYE	-	JOINED.	
DR	PFAM	PF00368	HMG-COA_red.1	-	JOINED.	
DR	PRINTS	PR00071	HMGCOARBASE	-	JOINED.	
DR	PROSITE	PS00066	HMG_COA_REDUCTASE.1	-	JOINED.	
DR	PROSITE	PS00318	HMG_COA_REDUCTASE.2	-	JOINED.	
DR	PROSITE	PS01192	HMG_COA_REDUCTASE.3	-	JOINED.	
DR	PROSITE	PS50065	HMG_COA_REDUCTASE.4	-	JOINED.	
KW	Cholesterol biosynthesis		Endoplasmic reticulum; Transmembrane;			
FT	DOMAIN	1	339	MEMBRANE-BOUND.		
FT	DOMAIN	340	449	LINKER.		
FT	DOMAIN	450	887	CATALYTIC.		
FT	TRANSMEM	10	39			
FT	TRANSMEM	57	78			
FT	TRANSMEM	90	114			
FT	TRANSMEM	124	149			
FT	TRANSMEM	160	187			
FT	TRANSMEM	192	220			
FT	TRANSMEM	315	339			
FT	ACT_SITE	558	558	BY SIMILARITY.		
FT	ACT_SITE	766	766	BY SIMILARITY.		
FT	ACT_SITE	865	865	GENERAL BASE (BY SIMILARITY).		
FT	CARBOHYD	281	281			
FT	CARBOHYD	517	517			
FT	CARBOHYD	869	869			
FT	VARIANT	733	733			
SEQ	SEQUENCE	887 AA;	97080 MW;	433IE53ADA250E6A CRC64;		
Query Match		2.0%;	Score 183;	DB 1;	Length 887;	
Best Local Similarity		18.7%;	Pred. 0.433e-12;			
Matches	31;	Conservative	68;	Mismatches 62;	Indels 5;	Gaps 5;
Db	61	DIITITRCIALIYIYFQFQNLROQLSKYILGIAGFTIFFSSFFVS-TYVIAH-LDKEL	118			
Oy	305	ELVPTVAFMLVFAVYFVSRKIDVFRSRLFLALCSVITTAGSLAMSIGCFEFGTLTSL	364			
Db	119	TGLNALPFFLLIDLSASALAFALSSNSODEYRENIARGMALIPPTLLDALVCLV	178			
Oy	365	QS-KNIFPYLIVLGLNSLIVITSVSMDETFVFKIRVAQALSKESGMHSLKTLTLEITI	423			
Db	179	IGCVTMSGVROLEIMCCGCMVLANFVEMTFPPACVSILVLELSR	224			
Oy	424	LTIGLATFVPYIQEFICFATVGLISDFMLQMLFST-I-LAMNIKR	467			
RESULT	12					
ID	HMDH_RAT	STANDARD;	PRT;	887	AA.	
AC	P51639;					
DT	01-OCT-1996	(Rel. 34, Created)				
DT	01-OCT-1996	(Rel. 34, Last sequence update)				
DT	15-DEC-1998	(Rel. 37, Last annotation update)				
DE	3-HYDROXY-3-METHYLGUTARYL-COENZYME A REDUCTASE (EC 1.1.1.34)	(HMG-COA REDUCTASE).				
GN	HMGCR.					
OS	Rattus norvegicus (Rat).					
OC	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
RN	[1]					
RP	SEQUENCE FROM N.A.					

RA Hangjong C., Williams D., Shapiro D.;
 RL Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP CHARACTERIZATION.
 RC TISSUE=LIVER.
 RX MEDLINE: 85261452.
 RA Edwards P.A., Kemper E.S., Ian S.-F., Erickson S.K.;
 RT "Functional size of rat hepatic 3-hydroxy-3-methylglutaryl coenzyme A
 RL reductase as determined by radiation inactivation.";
 RT J. Biol. Chem. 260:10278-10282(1985).
 RN [3]
 RP SEQUENCE OF 423-443 AND 667-678, AND PHOSPHORYLATION.
 RC TISSUE=LIVER.
 RX MEDLINE: 90316098.
 RA Clarke P.R., Hardie D.G.;
 RT "Regulation of HMG-CoA reductase: identification of the site
 RL phosphorylated by the AMP-activated protein kinase in vitro and in
 RT intact rat liver.";
 RL EMBO J. 9:2439-2446(1990).
 RN [4]
 RP PHOSPHORYLATION.
 RC TISSUE=LIVER.
 RX MEDLINE: 92331788.
 RA Gillespie J.G., Hardie D.G.;
 RT "Phosphorylation and inactivation of HMG-CoA reductase at the AMP-
 RL activated protein kinase site in response to fructose treatment of
 RT isolated rat hepatocytes.";
 RL FEBS Lett. 306:59-62(1992).
 CC -1- FUNCTION: THIS TRANSMEMBRANE GLYCOPROTEIN IS INVOLVED IN THE
 CC CONTROL OF CHOLESTEROL BIOSYNTHESIS. IT IS THE RATE-LIMITING
 CC ENZYME OF STEROL BIOSYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: (S)-MEVALONATE + COA + 2 NADP(+) = 3-HYDROXY-
 CC 3-METHYLGUTARYL-COA + 2 NADPH.
 CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC -1- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M29249; AAA40608.1;
 DR PFAM: PF00368; HMG-COA_red. 1.
 DR PRINTS: PR00071; HMGCOARDTASE.
 DR PROSITE: PS00066; HMG_COA_REDUCTASE_1; 1.
 DR PROSITE: PS00318; HMG_COA_REDUCTASE_2; 1.
 DR PROSITE: PS01192; HMG_COA_REDUCTASE_3; 1.
 DR PROSITE: PS50065; HMG_COA_REDUCTASE_4; 1.
 KW Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;
 KW Cholesterol biosynthesis; NADP; Phosphorylation.
 FT DOMAIN 1 339
 FT DOMAIN 340 449
 FT DOMAIN 450 887
 FT DOMAIN 10 39
 FT TRANSMEM 57 78
 FT TRANSMEM 90 114
 FT TRANSMEM 124 149
 FT TRANSMEM 160 187
 FT TRANSMEM 192 220
 FT TRANSMEM 313 339
 FT ACT_SITE 558 558
 FT ACT_SITE 766 766
 FT ACT_SITE 865 865
 FT CARBOHYD 281 281
 FT CARBOHYD 517 517
 FT CARBOHYD 869 869
 FT MOD_RES 871 871
 FT CATALYTIC.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT GENERAL BASE (BY SIMILARITY).
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT PHOSPHORYLATION.

FT CONFLICT 423 423 S -> T (IN REF. 3).
 FT CONFLICT 428 428 L -> V (IN REF. 3).
 FT CONFLICT 443 443 P -> N (IN REF. 3).
 SQ SEQUENCE 887 AA; 96592 MW; D941273D06573D22 CRC64;
 Query Match 2.0%; Score 183; DB 1; Length 887;
 Best Local Similarity 18.7%; Pred. No. 4.33e-12;
 Matches 31; Conservative 68; Mismatches 62; Indels 5; Gaps 5;
 Db 61 DIILITRCALIIYIPQFONLRQSGKYLIGAGLFTISSPFVS-TYVHF-LDKEL 118
 Qy 305 ELVPTVAFLVFAVYFSVKRIDFRSRFLALCSYTTTGSLAMSGLCFFELTSL 364
 Db 119 TGLNEALPFLLLIDSRASALAFALSSNSODEYRENIARGMALIGPTFLDALVECLV 178
 Qy 365 QS-KDIFVIVILVGLNLSLVTSSVSMDETFDKIVQAALSGEWHISKLTLETLI 423
 Db 179 IGVGTMSGVROLEIMCCGCGSVLANFVETTFPACSLVLELR 224
 Qy 424 LTIGLAFVPIVIOEFCIFALVGLSLDFMLMLFST-I-LAMNIRK 467
 RESULT 13
 ID HMDH_RABBIT STANDARD: PRT: 888 AA.
 AC Q29512;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE 3-HYDROXY-3-METHYLGUTARYL-COENZYME A REDUCTASE (EC 1.1.1.34) (HMG-COA
 DE REDUCTASE).
 GN HMGCR.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NEW ZEALAND WHITE; TISSUE=LIVER;
 RA Yamada M., Yoshimatsu M., Kinowaki M., Kai M., Setoguchi T.;
 RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: THIS TRANSMEMBRANE GLYCOPROTEIN IS INVOLVED IN THE
 CC CONTROL OF CHOLESTEROL BIOSYNTHESIS. IT IS THE RATE-LIMITING
 CC ENZYME OF STEROL BIOSYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: (R)-MEVALONATE + COA + 2 NADP(+) = (S)-3-
 CC 3-HYDROXY-3-METHYLGUTARYL-COA + 2 NADPH.
 CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM AND PEROXISOMES.
 CC -1- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.
 DR PFAM: PF00368; HMG-COA_red. 1.
 DR PRINTS: PR00071; HMGCOARDTASE.
 DR PROSITE: PS00066; HMG_COA_REDUCTASE_1; 1.
 DR PROSITE: PS00318; HMG_COA_REDUCTASE_2; 1.
 DR PROSITE: PS01192; HMG_COA_REDUCTASE_3; 1.
 DR PROSITE: PS50065; HMG_COA_REDUCTASE_4; 1.
 KW Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;
 KW Cholesterol biosynthesis; NADP; Peroxisome.
 FT DOMAIN 1 339
 FT DOMAIN 340 449
 FT DOMAIN 450 888
 FT DOMAIN 10 39
 FT TRANSMEM 57 78
 FT TRANSMEM 90 114
 FT TRANSMEM 124 149
 FT TRANSMEM 160 187
 FT TRANSMEM 192 220
 FT TRANSMEM 313 339
 FT ACT_SITE 559 559
 FT ACT_SITE 767 767
 FT ACT_SITE 866 866
 FT CARBOHYD 281 281
 FT CARBOHYD 296 296
 FT CARBOHYD 518 518
 FT CATALYTIC.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT GENERAL BASE (BY SIMILARITY).
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.

THIS PAGE BLANK (USPTO)

 W O R L D S I D E (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:34:50 2000; Maspar time 39.26 Seconds

Tabular output not generated. 897.114 Million cell updates/sec

Title: >US-09-332-522B-4
 Description: (1-508) from US09332522B.pep
 Perfect Score: 3806
 Sequence: 1 MDPFVFIVIASLYGVLYFF.....LIFALALIKVAMLSFLRPLL 508

Scoring table: PAM 150
 Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: sptrembl12

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
 5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
 9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
 13:sp_vertebrate 14:sp_virus

Statistics: Mean 50.285; Variance 102.588; scale 0.490

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	222	5.8	368	10	065444	1.15e-17
2	221	5.8	383	1	09YFP0	1.68e-17
3	212	5.6	377	1	057994	5.25e-16
4	193	5.1	389	1	027421	6.50e-13
5	175	4.6	386	1	P95972	4.51e-10
6	146	3.8	501	1	028947	9.58e-06
7	134	3.5	165	2	092868	4.91e-04
8	135	3.5	2846	5	09Y0A2	3.57e-04
9	135	3.5	2848	5	061193	3.57e-04
10	135	3.5	2857	5	09Y0A0	3.57e-04
11	135	3.5	2892	5	09Y0A1	3.57e-04
12	130	3.4	372	4	09Y8Y3	1.73e-03
13	130	3.4	376	4	043898	1.73e-03
14	125	3.3	4199	2	P74440	8.12e-02
15	122	3.2	325	13	042169	2.02e-02
16	120	3.2	344	8	078860	3.68e-02
17	123	3.2	715	5	09XUV0	1.49e-02
18	116	3.0	161	1	051981	1.20e-01
19	113	3.0	207	1	026302	2.85e-01
20	115	3.0	447	10	080885	1.50e-01

21	111	2.9	140	3	094594	5.04e-01
22	111	2.9	199	10	P93296	5.04e-01
23	112	2.9	244	5	018352	3.80e-01
24	112	2.9	326	13	042168	3.80e-01
25	112	2.9	487	2	033222	3.80e-01
26	109	2.9	589	5	09X85	8.86e-01
27	112	2.9	844	13	093552	3.80e-01
28	110	2.9	5138	10	092W94	6.69e-01
29	105	2.8	222	13	09W6V8	2.67e+00
30	107	2.8	232	1	09YCB9	1.54e+00
31	108	2.8	266	11	061198	1.17e+00
32	106	2.8	268	2	067364	2.03e+00
33	106	2.8	283	2	092905	2.03e+00
34	106	2.8	285	1	058347	2.03e+00
35	108	2.8	326	13	042609	1.17e+00
36	106	2.8	326	13	042166	2.03e+00
37	105	2.8	326	13	042167	2.67e+00
38	106	2.8	344	8	079641	2.03e+00
39	106	2.8	348	8	020949	2.03e+00
40	105	2.8	372	13	P79945	2.67e+00
41	108	2.8	373	1	058089	1.17e+00
42	106	2.8	470	4	043200	2.03e+00
43	107	2.8	770	2	P73845	1.54e+00
44	105	2.8	953	5	077046	2.67e+00
45	106	2.8	1157	5	P90747	2.03e+00

ALIGNMENTS

RESULT ID	1	PRELIMINARY:	PRT:	368 AA.
AC 065444:				
DT 01-AUG-1998 (TrEMBLrel. 07, Created)				
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)				
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)				
DE HYPOTHETICAL 40.3 KD PROTEIN.				
GN F1C12.220.				
OS Arabidopsis thaliana (Mouse-ear cress).				
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
OC [1]				
RP SEQUENCE FROM N.A.				
RA BEVAN M., TERRIN N., ARDILES W., BUYSHAERT C., DASSEVILLE R., DE CIERCK R., DE KEYSER A., MEYI P., ROUZE P., VAN DEN DAELHE H., VILLAROEI R., GIELEN J., VAN MONTAGU M., BANCROFT I., MEWES H.W., MAYER K., SCHUELLER C.,				
RU Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.				
RF [2]				
RP SEQUENCE FROM N.A.				
RA ED ARABIDOPSIS SEQUENCING PROJECT;				
RU Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.				
DR EMBL; AL022224; CAAL8255.1; -				
KW PRINTS: PRO1000; SREBPS2PTASE.				
KW Hypothetical protein.				
SO SEQUENCE 368 AA; 40315 MW; 036C342C CRC32;				
Query Match		5.8%;	Score 222;	DB 10; Length 368;
Best local similarity		35.3%;	Pred. No. 1.15e-17;	
Matches	48;	Conservative	34;	Mismatches 46; Indels 8; Gaps 8;
Db 53 FS-VLKWVMSIGIGFVASLIIVTVLLLOFHNPFSNRMTJAV-FGFS-PSTRVSL 109				
QY 78 FNVSVL-VTFSL-LPGLI-LTIATIFSSGEQDSSSVSPGVQLEILLPQVNP 134				
Db 110 GIAVAVSVTVSVHGLHAAASGIOMEXIAV-FIAIFGCVAFDNDVLSLPS 168				
QY 135 EIGYITTVLVCIVHMGHAAVMDVYTGIGINFICL-PLATLSHDLNLRW 193				
Db 169 FNALRYCAGIWMNAV 184				
QY 194 FRKLRLVLCAGIWMNFV 209				

Query	Subject	Score	Expect	Ident	Accession	Database	Length	Gaps
97	PMHPS-VOLVIRPCTVPLV-YGL-IALAL-VIVHELSGFARSDDPLKSVGLLFI	152						
115	PVGVVQVLEILLPVNPLERIGYITTVLVLVHEMHGALAAVMDVVTGEGILFI	174						
153	IIPAFVPEPDEDQKKAPLSRLVFCAGSEFANVALLISLLVNGIALAFEP	205						
175	CLPLATVETLSHDHLSRLRWPFRKLPLVLCAGIMHNVFVFGVGYLLISTYGI	227						
RESULT	4	PRELIMINARY;	PRT;	389	AA.			
ID	027421							
AC	027421							
DT	01-JAN-1998 (TReMBLrel. 05, Created)							
DT	01-JAN-1998 (TReMBLrel. 05, Last sequence update)							
DT	01-NOV-1999 (TReMBLrel. 12, Last annotation update)							
DE	CONSERVED PROTEIN.							
GN	MTF1368.							
OS	Methanobacterium thermoautotrophicum.							
OC	Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;							
OC	Methanobacterium.							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=DELTA H;							
RX	MEDLINE: 98037514.							
RA	SMITH D.R., DOUGETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,							
RA	ALDRIDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,							
RA	HARRISON D., HOANG L., KENGLE P., LUMM W., POTHIER B., QIU D.,							
RA	SPADAFORA R., VICARE R., WANG Y., WIERBOWSKI J., GIBSON R.,							
RA	JIYANI N., CARDOSO A., BUSH D., SAHER H., PATWELI D., PRABHAKAR S.,							
RA	MCODDALL S., SHIMER G., GOVAL A., PIETROWSKI S., CHURCH G.M.,							
RA	DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.,							
RT	"Complete genome sequence of Methanobacterium thermoautotrophicum							
RT	deltah: functional analysis and comparative genomics."							
RL	J. Bacteriol. 179:7135-7153(1997).							
DR	EMBL: AEO00889; AAB85845.1; -.							
DR	PFAM: PF00595; PDZ. 1.							
DR	PRINTS: PRO1000; SREBPS2PTASE.							
SO	SEQUENCE 389 AA; 42956 MW; DC436662 CRC32;							
Query Match	5.1%;	Score 193;	DB 1;	Length 389;				
Best Local Similarity	29.5%;	Pred. No. 6.50e-13;						
Matches 28;	Conservative 34;	Mismatches 32;	Indels 1;	Gaps 1;				
Db	113 LGYIGIGLATVYVHEFAFGILARLEGVRIKISIGLLAIIPGAFVPEPDEDIKIRPIS	172						
Oy	136 IGYITTVLVLVHEMHGALAAVMDVVTGEGILFIPLATVETLSHDHLSRLRWP	195						
Db	173 KMRIYAAGSVANLLIAGICFALFGISSFAMPAAF	207						
Oy	196 KLRVLCAGIMHNVFVFGVGY-LLISTYGITWSPly	229						
RESULT	5	PRELIMINARY;	PRT;	386	AA.			
ID	P95972							
AC	P95972							
DT	01-MAY-1997 (TReMBLrel. 03, Created)							
DT	01-MAY-1997 (TReMBLrel. 03, Last sequence update)							
DT	01-NOV-1999 (TReMBLrel. 12, Last annotation update)							
DE	ORF C04034.							
OS	Sulfolobus solfataricus.							
OC	Archaea; Crenarchaeota; Sulfolobales; Sulfolobus.							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=P2.							
RX	MEDLINE: 97055432.							
RA	SENSE S.C.W., KLENN H.P., SINGH R.K., ALLARD G., CHAN C.C.Y., LIU Q.Y.,							
RA	PENNY S.L., YOUNG F., SCHENK M.E., GAASTERLAND T., DOOLITTLE W.F.,							
RA	RAGAN M.A., CHARLEBOIS R.L.;							
RT	"Organizational characteristics and information content of an archaeal							
RT	genome: 156 kb of sequence from Sulfolobus solfataricus P2.";							
RL	MOL. Microbiol. 22:175-191(1996).							
DR	EMBL: Y08257; CAA69363.1; -.							


```
GN F33D4.2A.
OC Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditida;
OC Rhabditiina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BOERFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA JAMES M., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNEMMER E., STADEN R., STILSON J.,
RA THERRY-MEGE J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA JOHNSON D., STELLEYES L.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF036702; AAB88374.1;
DR PFAM: PF01365; RYDR-ITPR: 1.
DR PRINTS: PRO0779; INSP3RECEPT.
SQ SEQUENCE 2848 AA; 325857 MW; DC91ABD4 CRC32;

Query Match 3.5%; Score 135; DB 5; Length 2848;
Best Local Similarity 26.3%; Pred. No. 3.57e-04;
Matches 15; Conservative 26; Mismatches 14; Indels 2; Gaps 2;

Db 2466 LILASLCELLSSIGVTLT-LYIFGILQLVNKIVVAFAVNSKGLD-DRPIAEILAC 2520
Oy 208 FVFAGVCYLLSTVGTMSPLAYNQHVVAYVELTRKSPLEGRGLQVNDQITQVNGC 264

RESULT 10
ID O9Y0A0 PRELIMINARY; PRT: 2857 AA.
AC O9Y0A0;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE INOSITOL 1,4,5-TRISPHOSPHATE RECEPTOR.
DE ITR-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditida;
OC Rhabditiina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA BAYLIS H.A., FURUCHI T., YOSHIKAWA F., MIKOSHIBA K., SATTELLE D.B.;
RA "Inositol 1,4,5-trisphosphate receptors are strongly expressed in the
RA nervous system, pharynx, intestine and excretory cell of
RA Caenorhabditis elegans and are encoded by a single gene (itr-1).";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ243182; CAB45863.1;
KW Receptor.
SQ SEQUENCE 2857 AA; 327195 MW; CA316466 CRC32;

Query Match 3.5%; Score 135; DB 5; Length 2857;
Best Local Similarity 26.3%; Pred. No. 3.57e-04;
Matches 15; Conservative 26; Mismatches 14; Indels 2; Gaps 2;

Db 2475 LILASLCELLSSIGVTLT-LYIFGILQLVNKIVVAFAVNSKGLD-DRPIAEILAC 2529
```

```
Oy 208 FVFAGVCYLLSTVGTMSPLAYNQHVVAYVELTRKSPLEGRGLQVNDQITQVNGC 264

RESULT 11
ID O9Y0A1 PRELIMINARY; PRT: 2892 AA.
AC O9Y0A1;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE INOSITOL 1,4,5-TRISPHOSPHATE RECEPTOR.
DE ITR-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditida;
OC Rhabditiina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA BAYLIS H.A., FURUCHI T., YOSHIKAWA F., MIKOSHIBA K., SATTELLE D.B.;
RA "Inositol 1,4,5-trisphosphate receptors are strongly expressed in the
RA nervous system, pharynx, intestine and excretory cell of
RA Caenorhabditis elegans and are encoded by a single gene (itr-1).";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ243180; CAB45861.1;
KW Receptor.
SQ SEQUENCE 2892 AA; 330832 MW; 148D2E24 CRC32;

Query Match 3.5%; Score 135; DB 5; Length 2892;
Best Local Similarity 26.3%; Pred. No. 3.57e-04;
Matches 15; Conservative 26; Mismatches 14; Indels 2; Gaps 2;

Db 2510 LILASLCELLSSIGVTLT-LYIFGILQLVNKIVVAFAVNSKGLD-DRPIAEILAC 2564
Oy 208 FVFAGVCYLLSTVGTMSPLAYNQHVVAYVELTRKSPLEGRGLQVNDQITQVNGC 264

RESULT 12
ID O9Y5Y3 PRELIMINARY; PRT: 372 AA.
AC O9Y5Y3;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE ORPHAN G PROTEIN-COUPLED RECEPTOR GPR45.
DE GPR45.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 99156852.
RA MARCHESE A., SAWZDARGO M., NGUYEN T., CHENG R., HENG H.H., NOWAK T.,
RA IM D.S., LYNCH K.R., GORGE S.R., O'DOWD B.F.;
RA "Discovery of three novel orphan G-protein-coupled receptors."
RL Genomics 56:12-21(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC O'DOWD B.F.;
RA Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF118266; AAD21056.1;
KW Receptor.
SQ SEQUENCE 372 AA; 42000 MW; F8FA6D8E CRC32;

Query Match 3.4%; Score 130; DB 4; Length 372;
Best Local Similarity 27.4%; Pred. No. 1.73e-03;
Matches 32; Conservative 33; Mismatches 44; Indels 8; Gaps 8;

Db 19 SNADSGSTQLPAPRLSLAIVLMTYVGF-LGNTVVCITVGR-PAMRSAT-NLLIAT 75
Oy 108 SSSSVSPVGVPOLEILPVLNPLEIGIYITVLCVLVHMGHALAVMDVPTG 167

Db 76 LAFSDIM-LSLCCAPFTVTLITVRMHGDFHFCRLSATLYWF-FVLEBVALILLISV 130
Oy 168 FGKIFLCLPLATYELSHDHLNLSRW-F-RRLRYLCAGI-WHNEFVAGVCYLLSTV 221
```


THIS PAGE BLANK (USPTO)

Query Match 21.0%; Score 800; DB 1; length 510;
 Best Local Similarity 30.0%; Pred. No. 1,32e-138;
 Matches 155; Conservative 146; Mismatches 190; Indels 25; Gaps 19;

DB	SEQUENCE	510 AA	56498 MW	12B6B2257C0DA7A4 CRC64
FT	TRANSMEM	75	95	POTENTIAL.
FT	TRANSMEM	177	197	POTENTIAL.
FT	TRANSMEM	220	240	POTENTIAL.
FT	TRANSMEM	429	449	POTENTIAL.
FT	TRANSMEM	484	504	POTENTIAL.
FT	DOMAIN	109	127	POLY-SER.
FT	DOMAIN	371	375	POLY-SER.
SO	SEQUENCE	510 AA	56498 MW	12B6B2257C0DA7A4 CRC64

Query Match 21.0%; Score 800; DB 1; length 510;
 Best Local Similarity 30.0%; Pred. No. 1,32e-138;
 Matches 155; Conservative 146; Mismatches 190; Indels 25; Gaps 19;

DB	SEQUENCE	510 AA	56498 MW	12B6B2257C0DA7A4 CRC64
DB	1 MIPVSLVVVGGWVAVYVADLVKSSVYFKHSYEDMLEKNGISIPFIRMQTAVENRA 60			
OY	1 MPPVFFVILASVGLVYFFDFRFSKCMHY--PYDAFLKNTGISTINFMGLHMTSAFNRT 58			
DB	61 FYSWGRKRAKMLYQWFMFGVIMFSSFFLLGKTLMOQLAQMADSPSSSSSSSS 120			
OY	59 LTRWGSAGNSCT-RRYMIT-SENV-GVLVTFSLPIGLI--L-LTA--TIFSSGEDSSS 110			
DB	121 SSSSSSSINEOVLYVVGINLPVNOIYFFAVALISGVHEIGHGIAIREQVFNPF 180			
OY	111 SVSSPVGVV-Q-LEILLPGVNLPLEIGIYITTLVLCVHEHGHAAVMDVPTGFR 168			
DB	181 GI-FLFIIPGAFVDLFTTHLOLISPVQOLRIFCAGIMHNFVALIGIALVLLPYLLP 239			
OY	169 GIKFIFCL-PLATVETSHDLNLSRWFRKLRVLCAGIMHNFVAGCYLLISTVGTIMSP 227			
DB	240 FY-YTGVGLITEVAEDSPAIGRGLVGLDVLTHLQDCPVTVNODNECLDTATYEPQIG 298			
OY	228 LVAYNOH-VVTELTTRKSPLEGRGLQVNDQITQVNGCCPVNSESWATCLOH-SLKLK 285			
DB	299 YCISASTLOOLS-FPRPAKRLDOSTECNNHSLTDVCFEYRNPNFKRLTCTCPARKAVE 357			
OY	286 YCVSAFVOLNDESSAISHSIDGLOCCDELNPVNSCFEYVADAGDVPELPOHCLN 345			
DB	358 ANQ-VCRNKDKTS-SSSFCIVPSLETETRLIKVHP-QIDML---YVGHPLHLAY 410			
OY	346 VRTLEVESEHSSGVNCFCLPLRINTALMTFRKRONREKPLPVLYVGHVPMVTR 405			
DB	411 TVSITSEIPRFNPLSIDLPVIVETVFKYLLISGALAIYNAVPCFALDGMILNSFLDPT 470			
OY	406 IYEVSAFVPRYSLLKAMPDAMLLKLYNVFSGIALINAIAPFCGDGHAHITSTVISHP 465			
DB	471 LRSVIGDNVYKDLIGFILLGGSVLLAANTLGLWM 506			
OY	466 LVGRVDOHAKRDIISLITSVGLFALALLKVAWL 501			

RESULT 2
 MS2P_HUMAN STANDARD: PRT; 519 AA.

AC 043462;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE STEROL-REGULATORY ELEMENT-BINDING PROTEINS INTRAMEMBRANE PROTEASE
 DE (EC 3.4.24.-) (SITE-2 PROTEASE).
 GN 52P.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FIBROBLAST;
 RX MEDLINE: 98324087.
 RA Rawson R.B., Zelenski N.G., Nijhawan D., Ye J., Sakai J., Hasan M.T.,
 RA Chang T.Y., Brown M.S., Goldstein J.L.:
 RT "Complementation cloning of S2P, a gene encoding a putative
 RT metalloprotease required for intramembrane cleavage of SREBPs.";
 RL Mol. Cell 1:47-57(1997).
 -1- FUNCTION: INTRAMEMBRANE PROTEOLYSIS OF STEROL-REGULATORY ELEMENT-
 -1- BINDING PROTEINS (SREBPs) WITHIN THE FIRST TRANSMEMBRANE SEGMENT

CC THEREBY RELEASING THE N-TERMINAL SEGMENT WITH A PORTION OF THE
 CC TRANSMEMBRANE SEGMENT ATTACHED. SITE-2 CLEAVAGE COMES AFTER SITE-1
 CC CLEAVAGE WHICH TAKE PLACES IN THE LUMENAL LOOP.
 CC -1- COPACITOR: BINDS ONE ZINC ION (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG,
 CC LIVER, MUSCLE, KIDNEY AND PANCREAS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M50 (ZINC
 CC METALLOPROTEASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

CC EMBL: AF019612; AAC51937.1.-.
 CC PRINTS: PR01000; SREBPS2PTASE.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 CC Hydrolase; Metalloprotease; Zinc; Transmembrane.
 CC METAL 171 171 ZINC (CATALYTIC) (BY SIMILARITY).
 CC ACT_SITE 172 172 ZINC (CATALYTIC) (BY SIMILARITY).
 CC METAL 175 175 ZINC (CATALYTIC) (BY SIMILARITY).
 CC TRANSMEM 4 24 POTENTIAL.
 CC TRANSMEM 75 95 POTENTIAL.
 CC TRANSMEM 186 206 POTENTIAL.
 CC TRANSMEM 229 249 POTENTIAL.
 CC TRANSMEM 448 468 POTENTIAL.
 CC TRANSMEM 493 513 POTENTIAL.
 CC DOMAIN 109 136 POLY-SER.
 CC DOMAIN 380 384 POLY-SER.
 CC MUTAGEN 171 171 H->F: ACTIVITY ABOLISHED.
 CC MUTAGEN 172 172 E->A: ACTIVITY ABOLISHED.
 CC MUTAGEN 172 172 E->D: ACTIVITY PARTIALLY ABOLISHED.
 CC MUTAGEN 175 175 H->F: ACTIVITY ABOLISHED.
 CC SEQUENCE 519 AA; 57443 MW; 247D69E0FD747BD CRC64;

Query Match 19.0%; Score 723; DB 1; length 519;
 Best Local Similarity 29.9%; Pred. No. 6.10e-122;
 Matches 157; Conservative 146; Mismatches 188; Indels 34; Gaps 20;

DB	SEQUENCE	519 AA	57443 MW	247D69E0FD747BD CRC64
DB	1 MIPVSLVVVGGWVAVYVADLVKSSVYFKHSYEDMLEKNGISIPFIRMQTAVENRA 60			
OY	1 MPPVFFVILASVGLVYFFDFRFSKCMHY--PYDAFLKNTGISTINFMGLHMTSAFNRT 58			
DB	61 FYSWGRKRAKMLYQWFMFGVIMFSSFFLLGKTLMOQLAQMADSPSSSSSSSS 120			
OY	59 LTRWGSAGNSCT-RRYMIT-SENV-GVLVTFSLPIGLI--L-LTA--TIFSSGEDSS 108			
DB	121 SSSSSSSSSSSSSSLNHDOVLYVVGINLPVNOIYFFAVALISGVHEIGHGIAI 180			
OY	109 SSSVSPVGVV---V---Q-LEILLPGVNLPLEIGIYITTLVLCVHEHGHAAV 159			
DB	181 REQVRNGFGI-FLFIIPGAFVDLFTTHLOLISPVQOLRIFCAGIMHNFVALIGIAL 239			
OY	160 MEDVPTVGTGKIFLCL-PLATVETSHDLNLSRWFRKLRVLCAGIMHNFVAGCYLLI 218			
DB	240 VLLPYLLPYY-YTGVGLITEVAEDSPAIGRGLVGLDVLTHLQDCPVTVNODNECLD 298			
OY	219 STVGITMSPLYAYNOH-VVTELTTRKSPLEGRGLQVNDQITQVNGCCPVNSESWATCLO 277			
DB	299 TIAYEPQIGCIASASTLOOLS-FPRPAKRLDOSTECNNHSLTDVCFEYRNPNFKRLH 357			
OY	278 N-SLKLPGCVSADVQNLNDESSAISHSIDGLOCCDELNPVNSCFEYVADAGDV 336			
DB	358 CLPARKAVEATQ-VCRNKDKTS-SSSFCIVPSLETETRLIKVHP-QIDML---Y 410			
OY	337 ELPOHCLANRKLIEVESEHSSGVNCFCLPLRINTALMTFRKRONREKPLPVLY 396			
DB	411 VGHPLHLYTVTSITSEIPRFNPLSIDLPVIVETVFKYLLISGALAIYNAVPCFALDGM 470			


```

FT TRANSMEM 28 48 POTENTIAL.
FT TRANSMEM 52 72 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT TRANSMEM 216 236 POTENTIAL.
FT TRANSMEM 242 262 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 307 327 POTENTIAL.
FT TRANSMEM 341 361 POTENTIAL.
FT TRANSMEM 384 404 POTENTIAL.
SQ SEQUENCE 422 AA; 45809 MW; AF6570C0626D339 CRC64;

Query Match 3.0%; Score 113; DB 1; Length 422;
Best Local Similarity 23.9%; Pred. Nov. 7.24e-02;
Matches 22; Conservative 32; Mismatches 33; Indels 5; Gaps 5;

Db 96 LIIWISGIVNEFVFEHPLISLIGLFGAIVSATDPATLIPFISSMDIPDEVAITLAE 155
Oy 73 VMTSFNGLVATVFTSL-LPI-GLI-LLITFISSGQSDSSSVSPGVPLLEILLPGV 129
Db 156 SVFNDPLGI-VYTLI-CLSLALALAENITL 185
Oy 130 NLPEEIGYITTVLCVHEMGHALVAWE 161

RESULT 5 STANDARD; PRT; 450 AA.
AC Y325 HAEIN
AC P46440;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHETICAL PROTEIN HI0325.
GN HI0325.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE: 95350630.
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kertevs A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Goeyne J.D.,
RA Scott J.D., Shilley R., Liu L.-I., Glodex A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Sprilgus T., Hedblom E., Cotton M.D.,
RA Uteerback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrman J.L., Georgagis N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.
RT "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd.;"
RSL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: STRONG, TO V. PARAHAEMOLYTICUS HYPOTHETICAL PROTEIN IN
CC RNT 5 REGION (AC P46231).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL; U32717; AAC21988.1; -
DR TIGR; HI0325; -
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 53 73 POTENTIAL.
FT TRANSMEM 95 115 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.

```

```

FT FT TRANSMEM 148 168 POTENTIAL.
FT FT TRANSMEM 199 219 POTENTIAL.
FT FT TRANSMEM 242 262 POTENTIAL.
FT FT TRANSMEM 267 287 POTENTIAL.
FT FT TRANSMEM 302 322 POTENTIAL.
FT FT TRANSMEM 343 363 POTENTIAL.
FT FT TRANSMEM 378 398 POTENTIAL.
FT FT TRANSMEM 428 448 POTENTIAL.
SQ SEQUENCE 450 AA; 47439 MM; AFD139E73FDB061 CRC64;

Query Match 3.0%; Score 115; DB 1; Length 450;
Best Local Similarity 25.7%; Pred. No. 3,92e-02;
Matches 19; Conservative 24; Mismatches 28; Indels 3; Gaps 3;

Db 146 RRAVACVLTEGATATYLLPYGFGEKIFESTILVKNNINAGATLGGTAAOVSILAMLPV 205
Oy 71 RRAVMTSSNGVALYTFSLPIGL-ILLATITFSSEGDSSSVSPGVVP-VQLEIILPG 128
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Oy 129 VNLPLEEI-GYYIT 141
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 6 STANDARD; PRT; 320 AA.
ID AA3R_RAT
AC P28647; Q63792;
DT 01-DEC-1992 (Rel. 24, Created)
DD 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-FEB-2000 (Rel. 39, Last annotation update)
DN ADENOSINE A3 RECEPTOR (TGPR1).
GN ADORA3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE-BRAIN.
RX MEDLINE; 92366475.
RA Zhou Q.Y., Li C.Y., Olah M.E., Johnson R.A., Stiles G.L., Civelli O.;
RT "Molecular cloning and characterization of an adenosine receptor: the
RL A3 adenosine receptor.";
RN Proc. Natl. Acad. Sci. U.S.A. 89:7432-7436(1992).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR; TISSUE-TESTIS;
RX MEDLINE; 91285122.
RA Meyerhof W., Mueller-Brechlin R., Richter D.;
RT "Molecular cloning of a novel putative G-protein coupled receptor
RL expressed during rat spermiogenesis.";
RN FEBS Lett. 284:155-160(1991).
[3]
RN SEQUENCE FROM N.A.
RP TISSUE-BRAIN;
RX MEDLINE; 96196578.
RA Sajjad F.G., Boyle D.L., Domingo R.C., Firestein G.S.;
RT "cDNA cloning and characterization of Ai1, an alternatively spliced
RL rat A3 adenosine receptor variant.";
RN FEBS Lett. 382:125-129(1996).
CC -1- FUNCTION: RECEPTOR FOR ADENOSINE. THE ACTIVITY OF THIS RECEPTOR
IS MEDIATED BY G PROTEINS WHICH INHIBITS ADENYL CYCLASE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS ARE PRODUCED BY ALTERNATIVE
SPlicing.
CC -1- TISSUE SPECIFICITY: TESTIS, PARTICULARLY IN SPERMATOCTYES AND
SPERMATIDS BUT NOT IN SPERMAOOGONIA. LOW LEVELS IN THE BRAIN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING SPERMIOGENESIS.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```


CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X59249; CAA1937.1; -
DR EMBL: M94152; AAA40680.1; -
DR EMBL: X93219; CAA63702.1; -
DR PIR: S17177; S17177.
DR PIR: A46152; A46152.
DR HSSP: P29274; 1MMH.
DR GCRDB: GCR_0177; -
DR GCRDB: GCR_0373; -
DR GCRDB: GCR_1362; -
DR PFAM: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCR_RHODOPSN.
DR PRINTS: PR00424; ADENOSINER.
DR PRINTS: PR00555; ADENOSINER.
DR PROSITE: PS00237; G-PROTEIN_RECEPTOR; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Lipoprotein; Palmitate; Alternative splicing.
FT DOMAIN 1 16
FT TRANSSEM 17 39
FT DOMAIN 40 50
FT TRANSSEM 51 74
FT DOMAIN 75 86
FT TRANSSEM 87 108
FT DOMAIN 109 128
FT TRANSSEM 129 150
FT DOMAIN 151 179
FT TRANSSEM 180 200
FT DOMAIN 201 233
FT TRANSSEM 234 257
FT DOMAIN 258 263
FT TRANSSEM 264 286
FT DOMAIN 287 320
FT CARBOHYD 4 4
FT CARBOHYD 5 5
FT LIPID 85 168
FT LIPID 305 305
FT VARSPIC 119 119
FT CONFLICT 18 18
FT CONFLICT 68 68
FT CONFLICT 74 83
FT CONFLICT 132 132
FT CONFLICT 172 172
FT CONFLICT 288 288
FT CONFLICT 291 295
SQ SEQUENCE 320 AA; 35629 MW; 59DBA40C594ACDAB CRC64;
Query Match 2.9%; Score 111; DB 1; Length 320;
Best Local Similarity 27.4%; Pred. No. 1,33e-01;
Matches 32; Conservative 34; Mismatches 39; Indels 12; Gaps 10;
Db 51 YFVLSALADIAGVGLVPIPLAIAVSLVOMHFFYACLFMSCVLFTFHASIMSLAIAVDR 110
111 YLRKLVIRYRTVTTQRIMVFLGLCWL-VSFL-VGLTFMGMRKRVL-ELSONSS 164
138 YVITTLVLC-LVHHEMGHALA-AWVEDVPYTGFGIKFICFLPLAVTELSD-DHLN-SL-R 192
OY 193 WFR-KLEVLACGI--WHNVEFAGVCYLLISTVGTIMSPLYAVNOHVVTLLTRKSP 245
RESULT 7
ID NUSM-ALLMA STANDARD: PRT; 641 AA.
AC P50365;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN NDS OR NAD5.
OS Allomyces macrogynus.
OG Mitochondrion.
OC Eukaryota; Fungi; Chytridiomycota; Blastocladiiales; Blastocladiaceae;

CC Allomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 46923 / BURMA 3-35 (350C);
RA Paquin B., Roewer I., Wang Z., Lang B.F.;
RT "A robust fungal phylogeny using the mitochondrially encoded nads
RT protein sequence."
RL Can. J. Bot. 73:5180-5185(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 46923 / BURMA 3-35 (350C);
RA MEDLINE; 96226032.
RA Paquin B., Lang B.F.;
RT "The mitochondrial DNA of Allomyces macrogynus: the complete genomic
RT sequence from an ancestral fungus."
RL J. Mol. Biol. 255:688-701(1996).
CC -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U17010; AAB05846.1; -
DR EMBL: U41289; AAC49228.1; -
DR PFAM: PF00361; Oxidored_q1; 1.
DR PFAM: PF00662; Oxidored_q1_N; 1.
DR PFAM: PF00662; Oxidored_q1_N; 1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 641 AA; 70674 MW; 9C64C376B72E7E61 CRC64;
Query Match 2.9%; Score 109; DB 1; Length 641;
Best Local Similarity 29.4%; Pred. No. 2.41e-01;
Matches 15; Conservative 15; Mismatches 20; Indels 1; Gaps 1;
Db 141 GMEGIGSSYLLINFWFTRI-QANKASMOALVNRVGDPLGFLGFGIFWL 190
OY 451 GFDAHRTSYTHSFVLRVQVQHKRDLISLITISVGLFLALALIKYAWL 501
RESULT 8
ID YCF1 ARATH STANDARD: PRT; 1786 AA.
AC P56785;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 2.3.7 KDA PROTEIN YCF1.
GN YCF1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC Brassicaceae; Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Asamizu E., Tabata S.;
RT "Complete structure of the chloroplast genome of Arabidopsis
RT thaliana."
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC -I- FUNCTION: NOT YET KNOWN.
CC -I- SIMILARITY: TO A SIMILAR ORF IN OTHER PLANTS CHLOROPLAST.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CB.17;
RX MEDLINE: 95355830.
RA Stoneman E.R., Bennett M., An J., Chesnut K.A., Wakeford E.K.,
RT "Cloning and characterization of 5E6(ly-49c), a receptor molecule
RT expressed on a subset of murine natural killer cells.";
RL J. Exp. Med. 182:305-313(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=LUNG;
RX MEDLINE: 91332459.
RA Wong S., Freeman J.D., Kelleher C., Mager D., Takei F.;
RT "Ly-49 multigene family. New members of a superfamily of type II
RT membrane proteins with lectin-like domains.";
RL J. Immunol. 147:1417-1423(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6, AND 129/SVJ;
RX MEDLINE: 95356819.
RA Held W., Roland J., Raullet D.H.;
RT "Allelic exclusion of Ly49-family genes encoding class I MHC-specific
RT receptors on NK cells.";
RL Nature 376:355-358(1995).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A/SN;
RA Sundback J., Karre K., Sentman C.L.;
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=LUNG;
RA Wong S., Freeman J.D., Kelleher C., Mager D., Takei F.;
RL Submitted (JUN-1994) to the EMBL/Genbank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C, NZB, C57BL/6, AND BALB/C X C57BL/6;
RA Mathew P.A., Stoneman E., Bennett M., An J., Chesnut K.A.,
RL Wakeford E.K., Scheerer J.B., Siciliano M.J., Kumar V.;
CC Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: RECEPTOR ON NATURAL KILLER (NK) CELLS FOR CLASS I MHC.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U09739; AAA86873.1; -;
DR EMBL: U49865; AAA92951.1; -;
DR EMBL: U49866; AAA92952.1; -;
DR EMBL: U49867; AAA92953.1; -;
DR EMBL: U49868; AAA92954.1; -;
DR EMBL: U10305; AAA19053.1; -;
DR EMBL: U56405; AAB19101.1; -;
DR EMBL: U34891; AAA77066.1; -;
DR EMBL: U34892; G1003009; -;
DR HSSP: P04511; 1LIT.
DR MGD: MGI:101905; KLRK3.
DR PFAM: PF00059; lectin_c: 1.
DR PROSITE: PS00615; C-TYPE LECTIN 1; FALSE_NEG.
DR PROSITE: PS50041; C-TYPE LECTIN 2; 1
DR T-cell; Glycoprotein; Antigen; Transmembrane; Cell adhesion;
KW Signal-anchor; Lectin; Receptor; Multigene family;
FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 45 66 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).

FT DOMAIN 67 266
FT DOMAIN 143 261
FT DISULFID 154 257
FT DISULFID 171 249
FT CARBOHYD 79 79
FT CARBOHYD 87 87
FT CARBOHYD 104 104
FT CARBOHYD 113 113
FT CARBOHYD 160 160
FT VARIANT 2
FT VARIANT 22
FT VARIANT 34
FT VARIANT 42
FT VARIANT 60
FT VARIANT 66
FT VARIANT 72
FT VARIANT 85
FT VARIANT 93
FT VARIANT 115
FT VARIANT 117
FT VARIANT 127
FT VARIANT 129
FT VARIANT 131
FT VARIANT 133
FT VARIANT 146
FT VARIANT 151
FT VARIANT 174
FT VARIANT 179
FT VARIANT 189
FT VARIANT 198
FT VARIANT 219
FT VARIANT 226
FT VARIANT 232
FT VARIANT 247
FT VARIANT 251
FT VARIANT 260
SO SEQUENCE 266 AA; 31285 MM; 409F61E5DFA2299A CXC64;

Query Match 2.8%; Score 108; DB 1; Length 266;
Best Local Similarity 38.9%; Pred. No. 3, 25e-01;
Matches 14; Conservative 11; Mismatches 9; Indels 2; Gaps 2;

DB 39 CSAFWOLYKALGILICFLLYIVANVLAKEYNOH 74
QY 201 CAGIWHNFVFA-GV-CYLLISTVGTMTSPLYAYNOH 234

RESULT 12
ID CB2B_HUMAN STANDARD: PRT: 360 AA.
AC P34972;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)

[illegible]

```

DT 01-NOV-1997 (Rel. 35, last annotation update)
DE VASOACTIVE INTESTINAL POLYPEPTIDE RECEPTOR 1 PRECURSOR (VIP-R-1)
DE (PITUITARY ADENYLATE CYCLASE ACTIVATING POLYPEPTIDE TYPE II RECEPTOR)
DE (PACAP TYPE II RECEPTOR) (PACAP-R-2).
GN VIPRI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE; 92222309.
RA Ishihara T., Shigemoto R., Mori K., Takahashi K., Nagata S.;
RT "Functional expression and tissue distribution of a novel receptor
RT for vasoactive intestinal polypeptide."
RL Neuron 8:811-819(1992).
CC -1- FUNCTION: THIS IS A RECEPTOR FOR VIP. THE ACTIVITY OF THIS
CC RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLYL
CC CYCLASE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: IN LIVER, LUNG, INTESTINES, THYMUS AND BRAIN
CC (MOSTLY IN THE CEREBRAL CORTEX AND HIPPOCAMPUS).
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announcement@isb.ch).
CC or send an email to license@isb.ch.
CC -----
DR EMBL; M66835; AAA42331.1; -.
DR PIR; JH0594; JH0594.
DR GCRDB; GCR_0369; -.
DR PFM; PFO0002; 7tm_2; 1.
DR PRINTS; PR00249; GPCRSECRETIN.
DR PRINTS; PR00491; VASOACTIVEIPR.
DR PRINTS; PR01154; VIPRRECEPTOR.
DR PROSITE; PS00649; G-PROTEIN_RECPT_F2_1; 1.
DR PROSITE; PS00650; G-PROTEIN_RECPT_F2_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 30
FT CHAIN 31 459
FT FT
FT FT
FT DOMAIN 31 143
FT TRANSMEM 144 168
FT DOMAIN 169 175
FT TRANSMEM 176 195
FT DOMAIN 196 217
FT TRANSMEM 218 241
FT DOMAIN 242 255
FT TRANSMEM 256 277
FT DOMAIN 278 294
FT TRANSMEM 295 318
FT DOMAIN 319 343
FT TRANSMEM 344 363
FT DOMAIN 364 375
FT TRANSMEM 376 395
FT DOMAIN 396 459
FT CARBOHYD 58 58
FT CARBOHYD 69 69
FT CARBOHYD 100 100
FT CARBOHYD 292 292
FT SEQUENCE 459 AA; 52057 MW; 99E8957DA86698D2 CRC64;
Query Match 2.8%; Score 107; DB 1; Length 459;
Best Local Similarity 22.4%; Pred. No. 4,35e-01;
Matches 30; Conservative 40; Mismatches 54; Indels 10; Gaps 9

```

DB 283 DFGCWDITINSSLMWIKAPILSLVNFVLF-ICIRILVOKRPPDICKNDSSPYRSL 341
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
DB 342 AKSTLLIPLLEGH 355
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
QY 117 GVPVLELPGVN 130
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
RESULT 14
ID GTRL_MOUSE STANDARD: PRT: 492 AA.
AC P17809; Q61608;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSE TRANSPORTER TYPE 1, ERYTHROCYTE/BRAIN (Gtl1).
GN SLC2A1 OR GLUT1 OR GLUT-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 89240694.
RA Kaestner K.H., Christy R.J., McLeathlan J.C., Braiterman L.T.,
RA Cornelius P., Pekala P.H., Lane M.D.;
RT "Sequence, tissue distribution, and differential expression of mRNA
RT for a putative insulin-responsive glucose transporter in mouse 3T3-L1
RT adipocytes.";
RL Arch. Biochem. Biophys. 279:261-274(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90274408.
RA Reed B.C., Shade D., Alperovich F., Yang M.;
RT "3T3-L1 adipocyte glucose transporter (HepG2 class): sequence and
RT regulation of protein and mRNA expression by insulin,
RT differentiation, and glucose starvation.";
RL Arch. Biochem. Biophys. 279:261-274(1990).
RN [3]
RP SEQUENCE OF 151-237 FROM N.A.
RC STRAIN-C57BL/6; TISSUE-EMBRYO;
RX MEDLINE: 93170163.
RA Smith D.E., Gridley T.;
RT "Differential screening of a PCR-generated mouse embryo cDNA library:
RT glucose transporters are differentially expressed in early
RT postimplantation mouse embryos.";
RL Development 116:555-561(1992).
RN [4]
RP SEQUENCE OF 357-463 FROM N.A.
RX MEDLINE: 92111400.
RA Hogan A., Heyner S., Charron M.J., Copeland N.G., Gilbert D.J.,
RA Jenkins N.A., Thorens B., Schultze G.A.;
RT "Glucose transporter gene expression in early mouse embryos.";
RL Development 113:363-372(1991).
CC -!- FUNCTION: FACILITATIVE GLUCOSE TRANSPORTER. THIS ISOFORM MAY BE
CC RESPONSIBLE FOR CONSTITUTIVE OR BASAL GLUCOSE UPTAKE. HAS A VERY
CC BROAD SUBSTRATE SPECIFICITY: CAN TRANSPORT A WIDE RANGE OF ALDOSES
CC INCLUDING BOTH PENTOSE AND HEXOSE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- DEVELOPMENTAL STAGE: LEVELS DECLINE 3-FOLD BETWEEN DAYS 7.5 AND
CC 12.5 OF GESTATION. AT 7.5 DPC, EXPRESSED MORE STRONGLY IN
CC EXTREMEZYOTIC TISSUES THAN IN THE EMBRYO PROPER. EXPRESSED IN
CC AMNION, CHORION, AND ECTOPLACENTAL CONE. IN THE YOLK SAC,
CC EXPRESSED MORE STRONGLY IN THE MESODERM LAYER THAN THE ECTODERM.
CC EXPRESSION FAIRLY WIDESPREAD IN THE EMBRYO AT 8.5 DPC, BUT BY 10.5
CC DPC, EXPRESSION IS DOWN-REGULATED AND OBSERVED IN THE EYE AND THE
CC SPINAL CORD.
CC -!- SIMILARITY: HIGH TO OTHER GLUCOSE TRANSPORTERS.
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
DB 321 TVSLFVVERAGRRLH-LIGLAGAGCAVMTIALTERLPMWXYL 367
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
QY 460 TVHSFLVGRVDHAKRDIISLT-ITVGSLL-FRLALL-KVAMLSFL 504
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
RESULT 15
ID GTRL_BOVIN STANDARD: PRT: 492 AA.
AC P27674;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE GLUCOSE TRANSPORTER TYPE 1, ERYTHROCYTE/BRAIN.
GN SLC2A1 OR GLUT1 OR GLUT-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RA Bosdo R.J., Partridge W.M.;
RT "Molecular cloning of the bovine blood-brain barrier glucose
RT transporter cDNA and demonstration of phylogenetic conservation of the
RT 5' untranslated region.";
RL Mol. Cell. Neurosci. 1:224-232(1991).
Query Match 2.8%; Score 105; DB 1; Length 492;
Best Local Similarity 33.3%; Pred. No. 7.7e-01;
Matches 16; Conservative 19; Mismatches 9; Indels 4; Gaps 4;

ORGANISM	#formal_name Aeropyrum pernix
DATE	20-Aug-1989 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999.
ACCESSIONS	G727177
REFERENCE	A72450
authors	Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sakoi, Y.; Kikuchi, H. DNA Res. (1999) 6:83-101 Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.
#journal	
#title	Ctenarchaeon, Aeropyrum pernix K1.
#cross-references	MUID:99310338
#accession	G727177
#status	Preliminary
#molecule_type	DNA
#residues	1-383 #label KAW
#cross-references	DBJ:AP000058; NID:G5103388; PIDN:BA479121.1; PID:d1042897; PID:g5103600
#experimental_source	strain K1
GENETICS	
#gene	AP0209
SUMMARY	#length 383 #molecular-weight 39729 #checksum 8340
Query Match	5.8% Score 221 DB 2; Length 383;
Best Local Similarity	35.0% Pred. No.5.53e-16;
Matches	36; Conservative 27; Mismatches 39; Indels 1; Gaps 1;
Db	109 LIPGVTPMEDLVVAALGVVAHELGHAVVAAGVGRKNAGIALLFFPAFVELD 168 :: :: :: :: :: :: :: :: :: :: :: :: :: :: Oy 125 LLPEVNPLEIGIYITTLVLCVLVHMGMALAAMVEDPVGFGRKFICLPATLDS 184 :: :: :: :: :: :: :: :: :: :: :: :: :: :: Db 169 EEQIMKARLYSRKVFSAVTANIALLLT-LTIAMTAPVAEP 210 :: :: :: :: :: :: :: :: :: :: :: :: :: :: Oy 185 HDHLNSLRWFRKRLRVLCAGIMHFVEFGVCYLIIIVTGITMSP 227 :: :: :: :: :: :: :: :: :: :: :: :: :: ::
RESULT	3
ENTRY	A71450 #type complete
TITLE	hypothetical protein PH0256 - Pyrococcus horikoshii
ORGANISM	#formal_name Pyrococcus horikoshii
DATE	14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
ACCESSIONS	A71450
REFERENCE	A71000
authors	Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H. DNA Res. (1998) 5:55-76 Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3
#journal	
#title	Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3
#cross-references	MUID:98344137
#accession	A71450
#status	Preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type	DNA
#residues	1-377 #label KAW
#cross-references	GB:AP000001; NID:93236128; PID:d1030271; PID:93256645
#experimental_source	strain OT3
#note	this accession replaces an interim accession for a sequence replaced by Genbank
GENETICS	
#gene	PH0256
SUMMARY	#length 377 #molecular-weight 41845 #checksum 1083

Query Match	5.6%	Score 212	DB 2	Length 377
Best Local Similarity	32.7%	Pred. No. 1.32e-14		
Matches 37	Conservative	35	Mismatches 37	Indels 4
Gaps				4
Db	97	PVTMS-VOLVPGTPIV-YGL-IALALP-VIVHETSHFVARSBDIPKSGLLFI	152	
Qy	115	PGVGVQVLEILLPGVNLPLEEIGYITTLVCLVHEMGAALAVMDVPTVGIGIFIF	174	
Db	153	IIPGAFVPEDDQKKAPLRSRLVFAGSFAFNVALISILLVNGIALAFEP	205	
Qy	175	CLPIAYETLSHDHLSLRWFRKLRVLGAGIWNHVFVAGVCLLTSTVGITWSP	227	
RESULT	4			
ENTRY	H75213	#type complete		
TITLE	serine proteinase hira related protein PAB2195 - Pyrococcus abyssi (strain Orsay)			
ORGANISM	#formal_name Pyrococcus abyssi			
DATE	20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999			
ACCESSIONS	H75213			
REFERENCE	A75001			
#authors	anonymous, Genoscope			
#submission	submitted to the EMBL Data Library, July 1999			
#description	Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.			
#accession	H75213			
#status	preliminary			
#molecule_type	DNA			
#residues	1-378 #label KAM			
#cross-references	GB:AJ24883; GI:AO96836; NID:95457433; PIDN:CAB9159.1; PID:el515053; PID:95457668			
#experimental_source	strain Orsay			
GENETICS				
#gene	PAB2195			
SUMMARY	#length 378 #molecular_weight 41715 #checksum 4202			
Query Match	5.4%	Score 206	DB 2	Length 378
Best Local Similarity	30.2%	Pred. No. 1.07e-13		
Matches 49	Conservative	44	Mismatches 58	Indels 11
Gaps				9
Db	101	VPS-VQVIVPVTPIV-YGL-ISLAVL-IIVHETSHGFVARSBDIPKSGLLFI	156	
Qy	118	VPVQVLEILLPGVNLPLEEIGYITTLVCLVHEMGAALAVMDVPTVGIGIFIF	177	
Db	157	GAFVPEDDQKKAPLRSRLVFAGSFAFNVALISILLVNGIALAFEP	211	
Qy	178	LAYETLSHDHLSLRWFRKLRVLGAGIWNHVFVAGVCLLTSTVGITMSPLYAANQVVV	237	
Db	212	FGVYKSPAYGI--LEKGDVIVELNGVKINTLEEFIFMANT	251	
Qy	238	TELTRKSPLENGERGLQVNDITQVNGCPVNSESMVTCLONS	279	
RESULT	5			
ENTRY	H69048	#type complete		
TITLE	conserved hypothetical protein MTH1368 - Methanobacterium thermoautotrophicum (strain Delta H)			
ORGANISM	#formal_name Methanobacterium thermoautotrophicum			
DATE	05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998			
ACCESSIONS	H69048			
REFERENCE	A69000			
#authors	Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lum, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, A.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrkowski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.			
#Journal	J. Bacteriol. (1997) 179:7135-7155			

W P E R L (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:37:38 2000; MasPar time 14.63 Seconds
Tabular output not generated. 822.362 Million cell updates/sec

Title: >US-09-332-522B-4
Description: (1-508) from US09332522B.pep
Perfect Score: 3806
Sequence: 1 MDPFFVFIASLYGVLYFF.....LLFALLKVAWLSFLRPLL 508

Scoring table: PAM 150
Gap 11

Searched: 189963 seqs, 23696106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 36.607; Variance 170.723; scale 0.214

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	111	2.9	763	1 R11341	Human Thyroid Stimulat	8.63e+00
2	108	2.8	320	1 R79450	Rat A3 adenosine recep	1.37e+01
3	107	2.8	360	1 R67001	Cannabinoid receptor H	1.60e+01
4	107	2.8	459	1 R42848	VIP receptor protein.	1.60e+01
5	106	2.8	674	1 R12917	Human TSH Receptor wit	1.87e+01
6	106	2.8	714	1 R12915	Human TSH Receptor wit	1.87e+01
7	106	2.8	756	1 R12916	Human TSH Receptor wit	1.87e+01
8	106	2.8	764	1 R24244	Rat thyrotropin recept	1.37e+01
9	106	2.8	764	1 R12914	Human Thyroid Stimulat	1.87e+01
10	106	2.8	764	1 R12505	Human Thyroid Stimulat	1.87e+01
11	106	2.8	764	1 R13259	Human Thyroid Stimulat	1.87e+01
12	101	2.7	314	1 R27874	Odorant receptor clone	3.99e+01
13	103	2.7	336	1 R48733	G-protein coupled huma	2.95e+01
14	103	2.7	336	1 W02705	G-protein coupled huma	2.95e+01
15	103	2.7	428	1 W56671	Caenorhabditis elegans	2.95e+01
16	101	2.7	492	1 R11360	Glucose Transporter Pr	3.99e+01
17	98	2.6	216	1 R75910	Human olfactory recept	6.26e+01
18	98	2.6	216	1 R75909	Human olfactory recept	6.26e+01
19	99	2.6	312	1 R27873	Odorant receptor clone	5.39e+01
20	99	2.6	314	1 R27876	Odorant receptor clone	5.39e+01
21	98	2.6	365	1 W20791	H. pylori transmembran	6.26e+01
22	100	2.6	407	1 R04572	ORF4 product from the	4.64e+01
23	98	2.6	457	1 W69556	Rat equilibrative nucl	6.26e+01

24	100	2.6	492	1 W17835	Human glucose transpor	4.64e+01
25	99	2.6	492	1 W30000	Human GLUT1 protein.	5.39e+01
26	99	2.6	762	1 W55039	Homo sapiens TSH recep	5.39e+01
27	99	2.6	762	1 W14778	Human TSH receptor.	5.39e+01
28	98	2.6	2100	1 W89579	Calcium permeable volt	6.26e+01
29	98	2.6	2104	1 W57773	Musca domestica voltag	6.26e+01
30	98	2.6	2104	1 W89578	Calcium permeable volt	6.26e+01
31	98	2.6	2105	1 W89577	Calcium permeable volt	6.26e+01
32	98	2.6	2105	1 W57772	Musca domestica voltag	6.26e+01
33	97	2.5	263	1 R25424	Human MIP.	7.26e+01
34	95	2.5	284	1 W02718	G-protein coupled odor	9.75e+01
35	96	2.5	303	1 W32448	Mycobacterium tubercul	8.41e+01
36	96	2.5	303	1 W32380	Mycobacterium tubercul	8.41e+01
37	96	2.5	303	1 W81682	M. tuberculosis immuno	8.41e+01
38	96	2.5	303	1 W54320	Mycobacterium tubercul	8.41e+01
39	95	2.5	342	1 W81576	EBV-induced G-protein	9.75e+01
40	95	2.5	359	1 Y00890	Canine histamine H2 re	9.75e+01
41	95	2.5	365	1 W23606	Human P2Y4 receptor po	9.75e+01
42	96	2.5	514	1 R95044	Apoptosis participatin	9.75e+01
43	97	2.5	574	1 W81559	Respiratory syncytial	7.26e+01
44	97	2.5	574	1 R39285	Respiratory syncytial	7.26e+01
45	95	2.5	574	1 W47604	HRSV glycoprotein F.	9.75e+01

ALIGNMENTS

RESULT 1
ID R11341 standard; Protein; 763 AA.
AC R11341;
DT 05-JUN-1991 (first entry)
DE Human Thyroid Stimulating Hormone receptor.
KW thyroid stimulating hormone; TSH; receptor; Graves' Disease;
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..34
FT domain /label= signal peptide
FT domain 35..418
FT domain /label= extracellular domain
FT domain 419..440
FT domain /label= transmembrane domain I
FT domain 441..471
FT domain /label= transmembrane domain II
FT domain 495..517
FT domain /label= transmembrane domain III
FT domain 538..560
FT domain /label= transmembrane domain IV
FT domain 581..602
FT domain /label= transmembrane domain V
FT domain 625..648
FT domain /label= transmembrane domain VI
FT domain 660..681
FT domain /label= transmembrane domain VII
W09103483-A.
21-MAR-1991.
07-SEP-1990; U050866.
08-SEP-1989; US-404899.
10-AUG-1990; US-565669.
(NEWB-) NEW ENGLAND MED CEN.
Cone R;
WPI; 91-102023/14.
N-PSDB; Q11124.
Pure nucleic acid encoding thyroid stimulating hormone and
protein - used to detect increased levels of intracellular C-AMP
indicative of thyroid malfunction, and to treat Graves disease
and thyroid conditions
Claim 7; Fig 6; 3pp; English.
Total RNA was isolated from a human Graves' thyroid and a bovine
thyroid sample and separately reverse transcribed to give cDNA. Five
microlitres of the cDNA was used in a PCR with two degenerate
CC oligonucleotides having homology to transmembrane domains III and
CC IV of human Luteinizing Hormone receptor (see Q1121 and Q1122). The
CC amplified DNA was isolated, digested with HindIII and EcoRI and

RESULT 3
ID R67001 standard; Protein; 360 AA.

```

Query Match      2.8%; Score 107; DB 1; Length 459;
Best Local Similarity 22.4%;
Matches 30; Conservative 54; Mismatches 54; Indels 10; Gaps 9;

Db      228 MANF-FWLLVEGLY--LYTLVAVSFSESKRYF-WGVLIGCWGVPVFITL-WTVARIYE 282
Qv      1 MDPEVFVIVLASLYGVLY-FED-RFKECKMHPYDPAFLKNTGLSINFMSLHHWIT-SAFNR 57

```

Db	263	T	263
Qy	268	S	268
RESULT	7		
ENTRY	A69415	#type complete	
TITLE	conserved hypothetical protein AF1322 - Archaeoglobus fulgidus		
ORGANISM	#formal_name Archaeoglobus fulgidus		
DATE	05-Dec-1997	#sequence_revision 05-Dec-1997	#text_change
ACCESSIONS	A69415		
REFERENCE	A69250		
#authors	Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gunn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyriides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Uutterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.		
#journal	Nature (1997) 390:364-370		
#title	The complete genome sequence of the hyperthermophilic sulfate-reducing archaeon Archaeoglobus fulgidus.		
#cross-references	MUID:98049343		
#accession	A69415		
##status	preliminary; nucleic acid sequence not shown;		
	translation not shown		
##molecule_type	DNA		
##residues	1-501 ##label KLE		
##cross-references	GB:AE001012; GB:AE000782; NID:g2689335; PID:g2649354 TIGR:AF1322		
SUMMARY	#length 501 #molecular-weight 55698 #checksum 8125		
Query Match	3.8%; Score 146; DB 2; Length 501;		
Best Local Similarity	30.3%; Pred. No. 3.92e-05;		
Matches	36; Conservative 33; Mismatches 43; Indels 7; Gaps 6;		
Db	110	LLPGVN-PIPIWGTGTLVTVLHVESHAILCRVEGVTKSLCVIAL-IPGGFAE 167	
Qy	124	ILPGLNPLPEEGYITTLVLGVHMGHALAAYMEDVPVTGFGIKFICPLA-YTE 182	
Db	168	PERKEIMDKERTKSSARIFSGVSNFAVAFAIAFLFSLLPTQPALVAVNDGGVV 226	
Qy	183	LSH-DHLNSLR--WFKRLRVLCAGIWHNEVFAGVCY-LISIVGITMSPLYYNQHVV 237	
RESULT	8		
ENTRY	D72085	#type complete	
TITLE	hypothetical protein - Chlamydia pneumoniae (strain CWL029)		
ORGANISM	#formal_name Chlamydia pneumoniae		
DATE	23-Apr-1999	#sequence_revision 23-Apr-1999	#text_change
ACCESSIONS	D72085		
REFERENCE	A72000		
#authors	Ka-man, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Davis, R.W.; Stephens, R.S.		
#journal	Nature Genet. (1999) 21:385-389		
#title	Comparative genomes of Chlamydia pneumoniae and C. trachomatis.		
#cross-references	MUID:99206606		
#accession	D72085		
##status	preliminary		
##molecule_type	DNA		
##residues	1-165 ##label ARN		
##cross-references	GB:AE0011622; GB:AE001363; NID:c4375652; PID:g4375655		

[illegible]

```

Query Match          3.1%; Score 117; DB 1; Length 396;
Best Local Similarity 26.4%; Pred. No. 1.72e-01;
Matches 33; Conservative 42; Mismatches 41; Indels 9; Gaps 8;

Db 239 IAGSFANFATALLGGAGIGVFGKIGNOVASALVSTATAALLVCLALLPRANSE 298
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 75 ITSFNVGLVTFSLPLGLILLIATIFSS-GEQSSSSVSPGVP-VQLEILLPGVN-- 130
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 299 IHGLVLSIFWGIAAMITGLGMQVKVLAAPDADTDVAMALFS-G-IFNIGICAGALVGNQV 356
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 131 LPLEEIG-VY-IITLVCLVHVHENGHALAAMVEDVPTGFGIKFICLPPLAYTELSDHL 188
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 357 -SLHW 360
      |::|
QY 189 NSLRW 193

RESULT 12
ENTRY T08294 #type complete
TITLE hypothetical protein H0875 - Halobacterium sp. (strain NRC-1)
ORGANISM plasmid pNRC100
#variety strain NRC-1
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
ACCESSION T08294
REFERENCE Z16408
#authors Ng, W.V.; Ciuffo, S.A.; Smith, T.M.; Bumgarner, R.E.; Baskin, D.; Faust, J.; Hall, B.; Loretz, C.; Seto, J.; Slagel, J.; Hood, L.; Dassarma, S.

#journal Genome Res. (1998) 8:1131-1141
#title Snapshot of a large dynamic replicon in a halophilic Archaeon: megaplasmid or minichromosome?
#cross-references EMBL:AF016485; NID:g2822278; PID:g2822355
#experimental_source strain NRC-1

GENETICS
#genome plasmid pNRC100
SUMMARY #length 161 #molecular-weight 16413 #checksum 5265

Query Match          3.0%; Score 116; DB 2; Length 161;
Best Local Similarity 22.0%; Pred. No. 2.25e-01;
Matches 22; Conservative 28; Mismatches 47; Indels 3; Gaps 3;

Db 44 FVLSMGLAFLVAVGPAQAQSGVDVCDTGATGDLVFGAVAGLGLPATGFTYTKAG 103
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 86 FSLPLGLILLIATIFSSGQDSSSVSPGVPVQLEILLPGVN-LPLEEIGYIITLV 144
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 104 LSYWRAGNPEKNDAKEKLVMSCIGFGIV-TIALVSPEL 142
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 145 LC-LVVHENGHALAAMVEDVPTGFGIKFICLPPLAYTEL 183
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 13
ENTRY E69124 #type complete
TITLE cobalamin biosynthesis protein M related protein - Methanobacterium thermoautotrophicum (strain Delta H)
ORGANISM #formal_name Methanobacterium thermoautotrophicum
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
ACCESSION E69124
REFERENCE A69000
#authors Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Petrokovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

#journal J. Bacteriol. (1997) 179:7135-7155
#title Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.

#cross-references MUID:98037514
#accession E69124
#status preliminary: nucleic acid sequence not shown; translation not shown

##molecule_type DNA
##residues 1-207 #label MTH
##cross-references GB:AE000807; GB:AE000566; NID:g2621239; PID:g2621245
##experimental_source strain Delta H

GENETICS
#gene MTH200
#start_codon GTG
#start_codon GTG
CLASSIFICATION #superfamily cobalamin biosynthesis protein M
SUMMARY #length 207 #molecular-weight 22151 #checksum 3365

Query Match          3.0%; Score 113; DB 2; Length 207;
Best Local Similarity 31.6%; Pred. No. 5.00e-01;
Matches 25; Conservative 23; Mismatches 26; Indels 5; Gaps 5;

Db 19 VALINIGIFTYIFSKPKERRIVIGLFAAAAASVSVSPGVPVHF-FLIPLAAI 77
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 75 ITSFNVGLV-TFSLPLGILLIAT-IFSSGEQDSSS-SVSSSPGVPVQLEILLPGVNL 131
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 78 LLGPLTAVIVA-TLCIVIQ 95
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 132 PLEEIGYIITLVLCVLVH 150
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 14
ENTRY H64489 #type complete
TITLE hypothetical protein MJ1521 - Methanococcus jannaschii
ORGANISM #formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
ACCESSION H64489
REFERENCE A64300
#authors Buit, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.L.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

#journal Science (1996) 273:1058-1073
#title Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
#cross-references MUID:96337999
#accession H64489
#status preliminary: nucleic acid sequence not shown; translation not shown

##molecule_type DNA
##residues 1-422 #label BUL
##cross-references GB:U67593; GB:L77117; NID:g1592149; PID:g1500411; TIGR:MJ1521

GENETICS
#map_position FOR1498907-1500175
#start_codon GTG
CLASSIFICATION #superfamily hypothetical protein MJ0057
SUMMARY #length 422 #molecular-weight 45804 #checksum 5126

Query Match          3.0%; Score 115; DB 2; Length 422;
Best Local Similarity 23.9%; Pred. No. 2.94e-01;
Matches 22; Conservative 32; Mismatches 33; Indels 5; Gaps 5;

```

Search completed: Sun Sep 3 12:37:21 2000
Job time : 56 secs.

Claves disease and treatment of myriotoxicozosis

122 LEILPGVNPLEEIGYITTLVLCLVH-EMGHALAAM 1600

RESULT 8
 ID R24244 standard; Protein; 764 AA.
 AC R24244;
 DT 01-DEC-1992 (first entry)
 DE Rat thyrotropin receptor precursor.
 KW thyroid dysfunction; TSH; thyroid stimulating hormone;
 KW thyroid gland; FRTL-5 thyroid epithelial cell line; strain Fischer;
 KW transmembrane domain; Graves' Disease; hyperthyroidism.
 OS Rattus norvegicus.
 FH Key Location/Qualifiers
 FT peptide 1..21
 FT /label= signal_peptide
 FT protein 22..764
 FT /label= mature_TSH_receptor
 FT modified_site 77..79
 FT /label= N-glycosylation
 FT modified_site 99..101
 FT /label= N-glycosylation
 FT modified_site 177..179
 FT /label= N-glycosylation
 FT modified_site 198..200
 FT /label= N-glycosylation
 FT modified_site 302..304
 FT /label= N-glycosylation
 FT domain 1..400
 FT /label= extracellular
 FT domain 401..681
 FT /label= transmembrane
 FT domain 416..440
 FT /label= hydrophobic
 FT /note= "TM1"
 FT domain 451..473
 FT /label= hydrophobic
 FT /note= "TM2"
 FT domain 493..516
 FT /label= hydrophobic
 FT /note= "TM3"
 FT domain 538..560
 FT /label= hydrophobic
 FT /note= "TM4"
 FT domain 582..605
 FT /label= hydrophobic
 FT domain 626..648
 FT /label= hydrophobic
 FT /note= "TM5"
 FT domain 662..681
 FT /label= hydrophobic
 FT /note= "TM6"
 FT domain 748..750
 FT /label= hydrophobic
 FT /note= "TM7"
 FT modified_site 748..750
 FT /label= phosphorylation_site
 FT /note= "potential"
 FT
 PN W09208726-A.
 PD 29-MAY-1992.
 PE 15-NOV-1990; U06533.
 PR 15-NOV-1990; W0-U065339.
 PA (AKAM/) AKAMITU T.
 PA (BANT/) BAN T.
 PA (COHN/) COHN L D.
 PA (IKUY/) IKUYAMA S.
 PA (KOSU/) KOSUGI S.
 PA (SAJI/) SAJI M.
 PI Akamitu T, Ban T, Ikuyama S, Kohn LD, Kosugi S, Saji M;
 DR WPI: 92-200128/24.
 DR N-PSDB: Q25387.
 DR Nucleotide expressed from rat thyrotropin receptor gene - and
 PT corresp. polypeptides and monoclonal antibodies, useful for
 PT diagnosis and treatment of thyroid disorders, in gene therapy
 PS Disclosure: Fig 1: 44pp: English.
 CC The full-length rat thyrotropin receptor sequence was deduced from
 CC the coding sequence deposited in the Genbank data base under

CC accession number M34842. The protein has a mol. wt of ca. 87,000.
 CC The receptor can be used in assays for detecting ligands capable of
 CC binding the receptor, e.g. thyrotropin and anti-receptor
 CC antibodies. See also Q25383-Q25386 and R24243.
 SO Sequence 764 AA;
 Db 434 FVLPLTSHK-LTV-PRFLM-C-NLAFADEGCVLLILASVDLTHREYNHAI-DW 488
 Oy 4 FVEFLVSLGLVLFEDRFRKSCMHYPDAF-IKNTGLSTNEMSLMHTSAFRTLLRW 62
 Db 489 -QTGPGCNTAGFTVFA-SELVYTLVITDERVATIFA 526
 Oy 63 GSAGNSCTRWMTSFNVCVLVTFSLPLIGILLIATIFS 102
 RESULT 9
 ID R12914 standard; Protein; 764 AA.
 AC R12914;
 DT 25-SEP-1991 (first entry)
 DE Human Thyroid Stimulating Hormone Receptor.
 KW TSH; thyrotropin; grave's disease.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT domain 1..418
 FT /label= extracellular
 FT /note= "putative"
 FT domain 419..682
 FT /label= transmembrane
 FT /note= "putative"
 FT domain 683..764
 FT /label= short intracellular cytoplasmic
 FT /note= "putative"
 FT modified_site 99
 FT /label= N-glycosylation site
 FT modified_site 113
 FT /label= N-glycosylation site
 FT modified_site 177
 FT /label= N-glycosylation site
 FT modified_site 198
 FT /label= N-glycosylation site
 FT modified_site 302
 FT /label= N-glycosylation site
 PN W09109137-A.
 PD 27-JUN-1991.
 PE 19-DEC-1990; U07387.
 PR 20-DEC-1989; US-451973.
 PR 30-AUG-1990; US-575018.
 PA (RAPO/) RAPOPORT B.
 PI Rapoport B;
 DR WPI: 91-208163/28.
 DR N-PSDB: Q12529.
 DR Recombinant, enzymatically active thyrotropin receptor -
 PT generated in non-thyroidal eukaryotic cells, useful for study of
 PT Graves disease and treatment of thyrotoxicosis
 PS Claim 3; Fig 1: 160pp: English.
 CC A human thyroid cDNA library was screened with two synthetic probes
 CC based on the reported amino acid sequence of the 3rd and 4th
 CC transmembrane domains of a putative human thyrotropin receptor and
 CC related receptors. This sequence was deduced from the nucleotide
 CC sequence of a 4kb clone isolated from the library. The protein
 CC contains a putative signal peptide, 7 transmembrane domains (within
 CC the general transmembrane domain defined in the Features Table),
 CC 5 potential glycosylation sites and a very short intracytoplasmic
 CC region. See also Q12530-1 and R12915-7.
 SO Sequence 764 AA;
 Query Match 2.8%; Score 106; DB 1; Length 764;
 Best Local Similarity 25.0%; Pred. No. 1.87e+01;
 Matches 40; Conservative 39; Mismatches 69; Indels 12; Gaps 12;

Db	434	FVLLILLTSHRK-LNV-PRFLM-C-NLAFAFCMGMYLLIASVDLYTHSEYVNHAI-DW	488
Qy	4	FVFVFIVLSLVGVLYFFEDRFRKSCMHHPYDAF-LKNTGLSINFMSLHMHTSAFNRILLRW	62
Db	489	-QTGGCCTAGCFYVFA-SELSVTLTYITLERNVAIFAMRLDRKIRLRACAIMGW	546
Qy	63	GSAGSGCRRRMVITISFNVCGLVLTSLPTIGILLIATIFS-SGQDSSSVSPVGPVQ	121
Db	547	VCCFLAL-LPLVGISY-AKVSICLPMDTETPLAIV	584
Qy	122	LEILLPGVNLPLEEGYITTLVCLVNH-EMGHALAAM	160
RESULT	10		
ID	R12505	standard: Protein: 764 AA.	
AC	R12505:		
DT	10-SEP-1991	(first entry)	
DE	Human thyroid stimulating hormone.		
KW	TSH; thyrotropin; oocyte; anti-TSHr.		
OS	Homo sapiens.		
PN	EP-433509-A.		
PD	26-JUN-1991.		
PE	14-DEC-1989: 403493.		
PR	14-DEC-1989: EP-403493.		
PA	(HENN-) HENNING BERLIN GMBH.		
PI	Parmentier M, Libert F, Dumont J, Vassart G;		
DR	WPI: 91-186499/26.		
N-PSDB:	Q12164.		
PT	Polypeptide with thyrotropin receptor activity - used to		
PT	quantitatively detect thyrotropin or its antibodies.		
PS	Disclosure; Fig 8; 31pp; English.		
CC	The sequence is derived from a human genomic library by screening		
CC	with a canine probe.		
CC	The human TSHr may be used to quantitatively detect TSH or anti-TSH		
CC	recombinant Abs, and labeled Abs raised to the TSHr product may be		
CC	used to detect TSHr.		
Sequence	764 AA:		
Query Match	2.8%; Score 106; DB 1; Length 764;		
Best Local Similarity	25.0%; Pred. No. 1.87e+01;		
Matches	40; Conservative 39; Mismatches 69; Indels 12; Gaps 12.		
Db	434	FVLLILLTSHRK-LNV-PRFLM-C-NLAFAFCMGMYLLIASVDLYTHSEYVNHAI-DW	488
Qy	4	FVFVFIVLSLVGVLYFFEDRFRKSCMHHPYDAF-LKNTGLSINFMSLHMHTSAFNRILLRW	62
Db	489	-QTGGCCTAGCFYVFA-SELSVTLTYITLERNVAIFAMRLDRKIRLRACAIMGW	546
Qy	63	GSAGSGCRRRMVITISFNVCGLVLTSLPTIGILLIATIFS-SGQDSSSVSPVGPVQ	121
Db	547	VCCFLAL-LPLVGISY-AKVSICLPMDTETPLAIV	584
Qy	122	LEILLPGVNLPLEEGYITTLVCLVNH-EMGHALAAM	160
RESULT	11		
ID	R13269	standard: Protein: 764 AA.	
AC	R13269:		
DT	15-OCT-1991	(first entry)	
DE	Human Thyroid Stimulating Hormone Receptor.		
KW	TSH; thyrotropin; Ashimoto's Disease; Basedow's Disease.		
OS	Homo sapiens.		
FT	Key	Location/Ovalifiers	
FT	peptide	1..21	
FT	domain	/label= signal_peptide	
FT	modified_site	22..415	
FT	modified_site	77..79	
FT	modified_site	/label= N-linked glycosylation site	
FT	modified_site	/note= "putative"	
FT	modified_site	99..101	
FT	modified_site	/label= N-linked glycosylation site	

ID W56671 standard: protein; 428 AA.
 AC W56671;
 DT 11-SEP-1998 (first entry)
 DE Caenorhabditis elegans clk-1 protein.
 DE toc-1 protein; developmental rate; longevity; cellular physiology;
 KW cellular metabolism; cancer; clk-1 protein.
 OS Caenorhabditis elegans.
 PN W09817823-A1.
 PD 30-APR-1998.
 PF 17-OCT-1997; CA0768.
 PR 18-DEC-1996; US-033196.
 PR 21-OCT-1996; US-028977.
 PA (UYMC-) UNIV MCGILL.
 PI Barnes T, Ewbank J, Hekimi S, Lakowski B;
 DR WPI: 98-261516/23.
 PT New Caenorhabditis elegans clk-1 gene - used to obtain human clk-1
 sequence, useful for, e.g. cancer diagnosis
 PS Disclosure: page 28; 46pp; English.
 CC The present sequence represents the Caenorhabditis elegans toc-1
 CC protein. The toc-1 gene is present within the clk-1 operon of
 CC C.elegans. The invention claims for the C.elegans clk-1 protein
 CC (W56670) which is involved in the developmental rate and longevity
 CC at the cellular physiology level, where clk-1 mutants have a longer
 CC life and altered cellular metabolism relative to wild-type. The
 CC clk-1 gene may be cloned to identify related genes, for e.g the human
 CC clk-1 sequence can be identified and may be useful in the diagnosis
 CC and/or prognosis of cancer. The invention claims that downregulation
 CC of expression of clk-1 can be used to increase the life span of animals
 CC or humans. The invention also claims that if downregulation clk-1
 CC expression could be targeted to a particular tissue or organ, it could
 CC lead to a specific physiological slowing down of this tissue/organ and
 CC a concomitant slower rate of degradation by the ageing process.
 CC Alternatively, administration of an agent to promote tissue- or
 CC organ-specific overexpression of clk-1 could allow the physiological
 CC rates of tissues or organs to be increased, to treat pathological
 CC conditions causing a slowdown of physiological rate of tissues/organs
 CC in a patient.
 SQ Sequence 428 AA;

Query Match 2.7%; Score 103; DB 1; Length 428;

Best Local Similarity 43.9%; Pred. No. 2.95e+01;

Matches 17; Conservative 6; Mismatches 12; Indels 2; Gaps 2;

Db 211 INLCEHFKHDFAMADPYCCLLISVAFSTWPLSTY 247

OY 197 LRVLCAGIMHNFVAG-VCYLLIS-TWGITMSPLXAY 231

Search completed: Sun Sep 3 12:38:13 2000
 Job time : 35 secs.

THIS PAGE BLANK (USPTO)

cholesterol metabolism that are under the control of an associated sterol regulatory element-1 (SRE-1) enhancer sequence. SREBP proteins fall into two families, SREBP-1 and SREBP-2. Both proteins are members of a family of basic-helix-loop-helix-leucine zipper (bHLH-zip) transcription factors. Each have the ability to bind to SRE sequences and modulate SRE-mediated transcription. SRE-1 is a conditional enhancer found in the promoters for the low density lipoprotein (LDL) receptor and 3-hydroxy-3-methylglutaryl coenzyme A synthase genes. It increases transcription in the absence of sterols and is inactivated when sterols accumulate. Human SREBP-2 contains 1141 amino acids and has 4% identity with SREBP-1a. SREBP-1a was the first recognised member of this family and has 1147 amino acids. The resemblance between SREBP-1a and SREBP-2 includes an acidic N-terminus, a highly conserved bHLH-zip motif (71% identical), and an unusually long extension of 740 amino acids on the C-terminal side of the bHLH-zip region. SREBP-2 possesses one feature lacking in SREBP-1a, a Glu rich region (27% Glu over 121 residues). SREBP promotes SRE-1-mediated gene transcription, eg LDL receptor, production in the presence of sterols. SREBP identified in screening assays, may be used to reduce plasma cholesterol levels and in controlling hypercholesterolaemia and its associated diseases. Sequence 1141 AA;

[illegible]

Db	1011	QORDLGSLRLAHSF	PAYKVFLEHATVRLMAGGSP	1047
Qy	1031	QYAYNHYA-IINGTRGDCWRLFVLTCLRMINGANP	1066	
RESULT	2			
ID	R6390	standard; Protein; 1147 AA.		
AC	R6390;			
DT	04-AUG-1995	(first entry)		
DE	Human SREBP-1a.			
KW	Probe; sterol regulatory element binding protein			
KW	cholesterol metabolism; sterol regulatory element			
KW	SREBP-1; SREBP-2; basic-helix-loop-helix-leucine			
KW	transcription factor; low density lipoprotein;			
KW	3-hydroxy-3-methylglutaryl coenzyme A synthase;			
KW	hypercholesterolaemia; enhancer.			
OS	Homo sapiens.			
EH	Key	Location/Qualifiers		
FT	region	1..60		
FT	/note=	"Acidic region"		
FT	domain	61..177		
FT	/note=	"Pro/Ser rich domain"		
FT	region	324..394		
FT	/label=	BHLH-zip		
FT	region	324..333		
FT	/note=	"Basic region"		
FT	region	337..349		
FT	/note=	"Helix 1"		
FT	region	358..373		
FT	/note=	"Helix 2"		
FT	region	359..401		
FT	/note=	"Putative Leucine zipper"		
FT	domain	427..461		
FT	/note=	"Ser/Gly/Pro rich domain"		
FT	domain	462..1147		
FT	/note=	"C-terminal domain"		
PN	W094369922-A.			
PD	24-NOV-1994.			
PF	13-MAY-1994; U05300.			
PR	13-MAY-1993; US-061697.			
PR	01-OCT-1993; US-131365.			
PA	(TEXA.) UNIV TEXAS SYSTEM.			
PI	Briggs MR, Brown MS, Goldstein JL, Wang X;			
DR	WPI; 95-006813/01.			
DR	N-PSDB; Q79037.			
PT	New sterol regulator element binding protein -			
PT	prods. and screening assays for agents for reduc-			
PT	cholesterol levels (Eng)			
PS	Claim 8; Page 185-98; 305pp; English.			
CC	The sequences given in R6390-91 represent the s-			
CC	binding proteins (SREBP), SREBP-1a and SREBP-2.			
CC	proteins which are involved in the regulation of			
CC	cholesterol metabolism that are under the contr-			
CC	sterol regulatory element-1 (SRE-1) enhancer se-			
CC	fall into two families, SREBP-1 and SREBP-2. Bot			
CC	a family of basic-helix-loop-helix-leucine z			
CC	transcription factors. Each have the ability to			
CC	and modulate SRE-mediated transcription. SRE-1			
CC	found in the promoters for the low density lipop			
CC	and 3-hydroxy-3-methylglutaryl coenzyme A synthe			
CC	transcription in the absence of sterols and is			
CC	accumulate. Human SREBP-2 contains 1141 amino ac			
CC	with SREBP-1a. SREBP-1a was the first recogni			
CC	and has 1147 amino acids. The resemblance betwe			
CC	includes an acidic N-terminus, a highly conserve			
CC	identical), and an unusually long extension of			
CC	C-terminal side of the BHLH-zip region. SREBP-			
CC	lacking in SREBP-1a, a Glu rich region (27% Glu			
CC	SREBP promotes SRE-1-mediated gene transcrip			
CC	production in the presence of sterols. SREBP i			
CC	assays, may be used to reduce plasma cholest			
CC	controlling hypercholesterolaemia and its assoc			

```
SQ      Sequence      1147 AA;

Query Match      4.5%; Score 358; DB 1; Length 1147;
Best Local Similarity 25.9%; Pred. No. 6.19e-19;
Matches 195; Conservative 195; Mismatches 290; Indels 72; Gaps 54;

Db      87 LSGQAAAPSLSPPOAPTLKMYPSMPATSPGIGKEESVP-LSILOTTPQPLPGALL 145
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      115 LSG-EGFASMLNPLOTSP-PSGGYPADADAYRLSLAQLAAPAMTPHOAASLFWNTNGID 172
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      146 POSFP-AP-APPFSSFPVLGYSPGPGFGSTGSPGNTQOPLPLASPPGVPPVLSHT 203
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      173 QKNFTHAMLSPPHTSMTSQPYTEAMCHINGMSYDQAQGPSYYSQHQSPPPHH 232
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      204 QVGSVVP-QOLLTVTAAPTAAPVTVTTSQIQVVPVLLQPHFIKADSL-LTAMK-TDGA 260
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      233 HHHPMPRIHENPEQVASFSDIEDAPETKPTHLVEPQSPKSPQNKKELLRLLVNMPSEVE 292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      261 TVKA--AGLSPLYSGTTVOTGPTLVSGGTILATVPLVDAEKLPIINRLAAGSKAPASA 318
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      293 RLKNNKSGACSATNGPSRSKEKAKIVIOETAEGDED-EDDEDSGETMSQGTIIIVR- 350
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      319 QSRGEKTAHNAIEKRYRSSINDKIIELKDLVGTAEAKLNKSAVLRAIDYIRFLQHSNQ 378
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      351 RPXTERTAHNLIEKRYCSINDRIQLKVLCCGDEAKLSKATLRAIERHIEVEHENQ 410
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      379 KLKOENLSLTAVHKSLSKDLVSACSGGNTDVLMEGVKTEVEDTLTPPPSDAGSPFQS 438
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      411 VLKHHVQMRKTLQNNR-L-PY-PE--PIQYTE-Y--SARSPVESPPPRNE---R-KR 458
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      439 SPLSLGSRGSGSGSDSPDSVFEDSKAKPEQPSLSHRGMLDRSLALCTLVFLCL 498
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      459 SRMS-TT-TPMKNGT-RDGSCKVTLF--AMLLAVLI--FNPIGLLA-G-S--A-IF-- 502
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      499 SCNPLASLLGARGLPSPDSTTSVYHSPGRNVLGTESRDGPGCAQWLLP-PVWLLNGLLV 557
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      503 S-KA-AA--EAP-IASFEGHVRIDDP--D--GTSTT-LFWECSIINMSYVWFNLMI 552
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      558 LVSLVLEFVGEVTPRPHSGPAYFWRRKQADLDLARGDFAQAQQLWLALRALJGRPLP 617
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      553 IYVVKLLIHGDPVQDFPMSVSWQTFVTREKARAEKNSGLNKAQKRFCECLATLDRSLP 612
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      618 TSHLDIACSLWNLRHLLOBLWGRWLAGRAGLQODCALRVDSASARDAALVYHKLH 677
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      613 SPGVDSVFSVGWECVREHLLNWLIGRYIARRRSTTKPVSV-V-CRSHAQTAVL-YHEIH 669
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      678 QLHTMG-K-H--T-G-GHLTATNALISALNLAECAG--DAVSATLAETIYVAAALRVK 727
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      670 QLHLMGITGFEDITYEFSALTGLFMSLCVNLAEAGASNDGLPRVYMAQIYISASIQCR 729
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      728 TSLPRAL-HFLTRFLSSAR-QACLAQSGSVPPAMQWLCHPVGRHFFVDGDW-S-VLSTP 783
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      730 LALPNLLAPFSGYFLRRARRHVRRAPEHVSVHLL-WIEHPATRKFSDAKRLHVLSSK 788
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      784 WESL-Y-SLAGN-PVDPLAQVTLFREHLLER 812
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      789 QKQLRFGSFEDEQLSPLARITRTLKVYLLSK 820
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT      3
ID      Y07242 standard; protein; 1829 AA.
AC      Y07242;
DT      06-JUL-1999 (first entry)
DE      Actin-filament binding protein 1-Afadin.
KW      Actin-binding protein 1-Afadin; rat; foetal brain tissue; infiltration;
KW      metastasis; carcinoma; diagnosis.
OS      Rattus sp.
PN      EP-905239-A2.
PD      31-MAR-1999.
PF      21-SEP-1998; 307643.
PR      22-SEP-1997; JP-257043.
PA      (NISC-) JAPAN SCI & TECHNOLOGY CORP.
CC      (OBAI/) OBAISHI H.
```

```
PA      (WADA/) WADA M
DR      WPI; 99-206773/18.
PT      New actin filament-binding protein 1-Afadin is useful for diagnosing
PT      and treating carcinomas
PS      Claim 1; Page 8-18; 27pp; English.
CC      This sequence represent an actin-binding protein 1-Afadin derived from
CC      rat foetal brain tissue. The isolated protein has a molecular weight
CC      of around 205 KD. 1-Afadin and its derivatives may be used to clarify
CC      the mechanisms of infiltration and metastasis of carcinoma useful in
CC      diagnosing or treating carcinomas.
SQ      Sequence 1829 AA;

Query Match      1.9%; Score 151; DB 1; Length 1829;
Best Local Similarity 23.4%; Pred. No. 3.08e-02;
Matches 59; Conservative 68; Mismatches 110; Indels 15; Gaps 14;

Db      1476 QRPQETVIRELOPQOQPRTIERRDLOVITTSKEELSGDLSL-PDPWKRDAAREKLEKQQ 1534
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      222 QHQSPPPHHHHHPMKIHENPE-QVASPSIEDAPETKPTHLVEPOSPPQNM-KEEL 279
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1535 MHI-VDMLSKREIHELQNGDRTAEESDRLRKLMLEWQFKRLQESKQKDEDEDEDDV 1593
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      280 LRLLVNMPSEVERLKNKSGACSATNGPSRSK-EKAAKIVIOETAEGDEDEDESDSG 338
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1594 DTMLIMORLAEARRLQDEERRRQOOLEMRKR-EVEDRVRQDEEDGRHQEERVKRDAAE 1652
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      339 ETMSQGTIIIV-RRPK--TE-RRTAHNLIEKKYRCSINDRIQQLKVLICGDEAKLSKAT 394
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1653 EKRRQEGSYSLRAEARRRROHEEAARLLPEEPGLSRPPLPODY-EPPSOSSAPSPPP 1711
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      395 LRAIEHIEVHEHNVQVKKHVEQMRTLQ-NNR-LPYPE-PIQYTEYSARSPVESPPSP 451
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1712 PP-QRNASYLKT 1722
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      452 PRNERKPSRMST 463
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT      4
ID      W52821 standard; Protein; 597 AA.
AC      W52821;
DT      08-SEP-1998 (first entry)
DE      Human PRCC-TFE3 construct protein from cell lines UOK120 and UOK146.
KW      PRCC; papillary renal cell carcinoma; TFE3; transcription factor;
KW      fusion protein; translocation; diagnosis; treatment.
OS      Homo sapiens.
FH      Key
FT      Protein
FT      1. .597
FT      Location/Qualifiers
FT      /label= PRCC-TFE3
FT      /note= "Fusion protein"
FT      Misc_difference 1. .156
FT      /label= PRCC
FT      /note= "papilloma renal cell virus partial sequence"
FT      Misc_difference 159. .597
FT      /label= TFE3
FT      /note= "transcription factor partial sequence"
PN      WO9806871-A1.
PD      19-FEB-1998.
PF      13-AUG-1997; G02209.
PR      13-AUG-1996; GB-016986.
PA      (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PI      Clark J, Cooper C, Shipley J;
DR      WPI; 98-159557/14.
DR      N-PSDB; V20957.
PT      Diagnosing papillary renal cell carcinoma by detecting gene
PT      trans-location - resulting in fusion of TFE3 gene with some other
PT      gene, also related vectors, transformed cells, specific binding
PT      reagents, peptide(s) encoded by fusions and therapeutic anti-sense
PT      sequences
PS      Claim 9; Fig 2b; 7lpp; English.
CC      This sequence represents a novel fusion protein constructed from a
CC      papillary renal cell carcinoma (PRCC) associated protein and the
CC      transcription factor TFE3 which is found in cell lines UOK120 and UOK146
```

CC and used in a method for the diagnosis, prophylactic and therapeutic
 CC treatment of papillary renal cell carcinoma. The translocation t(X;1)
 CC (p11.2;q21.2) found in PRCC results in a fusion of the TFE3 gene with a
 CC new chromosome 1 gene designated PRCC (at 1q21.2), resulting in
 CC expression of a fusion protein between the N-terminus of PRCC and almost
 CC the whole of the TFE3 gene. Normal TFE3 transcripts are no longer
 CC produced. Two other fusion partners for TFE3 have also been detected;
 CC NonO, from a invX (p11.2; q13-24 or 12) translocation and the PSF splice
 CC factor gene, resulting in t(X;1) (p11.2;p34). These trans-locations
 CC define a subgroup of PRCC generally encountered in patients younger
 CC than 25.
 CC Synthetic.
 SQ Sequence 597 AA;

Query Match 1.8%; Score 143; DB 1; Length 597;

Best Local Similarity 33.8%; Pred. No. 1.17e-01;

Matches 26; Conservative 23; Mismatches 23; Indels 5; Gaps 4;

Db 325 QKDNHNLIRRRNFNINDRKELGTLPKSSDPENRWNGTILKASVDYIRKLQKQQR 384

QY 355 ERTAHNLIEKKYRCSINDRIQLKVL--CGD-EAKLSKATLRAIEHIEVEHENQV 411

Db 385 SKD-LESQRSLEQANR 400

QY 412 LKHVEQMRKTL-QNNR 427

RESULT 5

ID W52827 standard; Protein; 618 AA.

AC W52827; 1998 (first entry)

DE Human chromosome X TFE3 protein.

KW PRCC; papillary renal cell carcinoma; TFE3; transcription factor;

KW fusion protein; translocation; diagnosis; treatment; chromosome X.

OS Homo sapiens.

PN W09806871-A1.

PD 19-FEB-1998.

PF 13-AUG-1997; G02209.

PR 13-AUG-1996; GB-016986.

PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.

PI Clark J, Cooper C, Shipley J;

DR WPI; 98-159557/14.

DR N-PSDB; V20963.

PT Diagnosing papillary renal cell carcinoma by detecting gene

PT trans-location - resulting in fusion of TFE3 gene with some other

PT gene, also related vectors, transformed cells, specific binding

PT reagents, peptide(s) encoded by fusions and therapeutic anti-sense

PT sequences

PS Claim 13; Fig 5B; 7lpp; English.

CC This sequence represents the human transcription factor TFE3 which is
 CC used in a method for the diagnosis, prophylactic and therapeutic
 CC treatment of papillary renal cell carcinoma (PRCC). The translocation
 CC t(X;1) (p11.2;q21.2) found in PRCC results in a fusion of the
 CC transcription factor, TFE3 gene with a new chromosome 1 gene designated
 CC PRCC (at 1q21.2), resulting in expression of a fusion protein between the
 CC N-terminus of PRCC and almost the whole of the TFE3 gene. Normal TFE3
 CC transcripts are no longer produced. Two other fusion partners for TFE3
 CC have also been detected; NonO, from a invX (p11.2; q13-24 or 12)
 CC translocation and the PSF splice factor gene, resulting in t(X;1)
 CC (p11.2;p34). These trans-locations define a subgroup of PRCC generally
 CC encountered in patients younger than 25.
 CC Synthetic

Query Match 1.8%; Score 143; DB 1; Length 618;

Best Local Similarity 33.8%; Pred. No. 1.17e-01;

Matches 26; Conservative 23; Mismatches 23; Indels 5; Gaps 4;

Db 346 QKDNHNLIRRRNFNINDRKELGTLPKSSDPENRWNGTILKASVDYIRKLQKQQR 405

QY 355 ERTAHNLIEKKYRCSINDRIQLKVL--CGD-EAKLSKATLRAIEHIEVEHENQV 411

Db 406 SKD-LESQRSLEQANR 421

QY 412 LKHVEQMRKTL-QNNR 427

RESULT 6

ID W52822 standard; Protein; 700 AA.

AC W52822; 1998 (first entry)

DE Human NonO/TFE3 fusion product from cell line UOK109.

KW PRCC; papillary renal cell carcinoma; TFE3; transcription factor;

KW fusion protein; translocation; diagnosis; treatment; NonO; p54-nrb.

OS Homo sapiens.

OS Synthetic.

FH Location/Qualifiers

FT Key

FT Protein

FT /label= NonO/TFE3

FT /note= "Fusion protein"

FT Misc_difference 1..377

FT /label= NonO

FT /note= "also known as p54-nrb"

FT Misc_difference 378..700

FT /label= TFE3

FT /note= "transcription factor sequence"

PN W09806871-A1.

PD 19-FEB-1998.

PF 13-AUG-1997; G02209.

PR 13-AUG-1996; GB-016986.

PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.

PI Clark J, Cooper C, Shipley J;

DR WPI; 98-159557/14.

DR N-PSDB; V20958.

PT Diagnosing papillary renal cell carcinoma by detecting gene

PT trans-location - resulting in fusion of TFE3 gene with some other

PT gene, also related vectors, transformed cells, specific binding

PT reagents, peptide(s) encoded by fusions and therapeutic anti-sense

PT sequences

PS Claim 9; Fig 3A; 7lpp; English.

CC This sequence represents a novel fusion protein constructed from the NonO
 CC protein (also known as p54-nrb) and the transcription factor TFE3 which
 CC is found in cell line UOK109 and used in a method for the diagnosis,
 CC prophylactic and therapeutic treatment of papillary renal cell carcinoma.
 CC The translocation t(X;1) (p11.2;q21.2) found in papillary renal cell
 CC carcinoma (PRCC) associated protein (PRCC) results in a fusion of the
 CC TFE3 gene with a new chromosome 1 gene designated PRCC (at 1q21.2),
 CC resulting in expression of a fusion protein between the N-terminus of
 CC PRCC and almost the whole of the TFE3 gene. Normal TFE3 transcripts are
 CC no longer produced. Two other fusion partners for TFE3 have also been
 CC detected; NonO, from a invX (p11.2; q13-24 or 12) translocation and the
 CC PSF splice factor gene, resulting in t(X;1) (p11.2;p34). These
 CC trans-locations define a subgroup of PRCC generally encountered in
 CC patients younger than 25.
 CC Synthetic

Query Match 1.8%; Score 143; DB 1; Length 700;

Best Local Similarity 33.8%; Pred. No. 1.17e-01;

Matches 26; Conservative 23; Mismatches 23; Indels 5; Gaps 4;

Db 428 QKDNHNLIRRRNFNINDRKELGTLPKSSDPENRWNGTILKASVDYIRKLQKQQR 487

QY 355 ERTAHNLIEKKYRCSINDRIQLKVL--CGD-EAKLSKATLRAIEHIEVEHENQV 411

Db 488 SKD-LESQRSLEQANR 503

QY 412 LKHVEQMRKTL-QNNR 427

RESULT 7

ID W52820 standard; Protein; 834 AA.

AC W52820;

DE Human PRCC-TFE3 fusion protein from cell line UOK124.

KW PRCC; papillary renal cell carcinoma; TFE3; transcription factor;

KW fusion protein; translocation; diagnosis; treatment.

OS Homo sapiens.

OS Synthetic.


```

PR 01-JUN-1994; US-252966.
PR 19-SEP-1991; US-756195.
PR 23-JUN-1992; US-903710.
PR 01-APR-1994; US-222638.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI Ayer DE, Eisenman RN;
DR WPI: 97-258216/23.
DR N-PSDB: T70134.
PT msin nucleic acids encoding recombinant polypeptide(s) that
PT associate with Mad polypeptide - are possible homologues of S.
PT cerevisiae general repressor protein
PS Example 2; Fig 2D; 11lpp; English.
CC This sequence represents Max 14, a basic helix-loop-helix z4pper
CC (bHLH2ip) protein. Max 11 and Max 14 cDNAs appear to be partial.
CC from a Manca (human Burkitt's lymphoma cell line) lambda gt10 library
CC overlapping cDNAs. Subsequent isolation of several overlapping cDNAs
CC permitted deduction of an apparently complete open reading frame for
CC Max that encodes 151 residues. The 9-amino acid insertion found in
CC several PCR clones is not indicated in the specification. Max is an
CC obligate partner for the DNA binding and transcriptional functions of
CC Myc family proteins as well as for the Mad protein. Max is a stable,
CC ubiquitously expressed protein which in general does not appear to be
CC regulated during mitogenesis, the cell cycle, or differentiation.
CC Expression of Mad is closely linked to differentiation in at least two
CC distinct cell lineages. The switch from Myc:Max to Mad:Max complexes may
CC reflect the repression of transcription of Myc regulated genes by Mad.
CC The DNA, vectors and host cells of the invention are useful for the
CC recombinant production of msin proteins useful in elucidation of Mad
CC repressor functions.
SQ Sequence 151 AA;

Query Match 1.6%; Score 125; DB 1; Length 151;
Best Local Similarity 24.4%; Pred. No. 2.16e+00;
Matches 20; Conservative 29; Mismatches 31; Indels 2; Gaps 2;

Db 1 MSDNDIDIEV-SDAKRAHNALEKRRDHDKDSFSLRDSVPSLQGEKASRAQILDKAT 59
QY 341 MSQGTIIIVRRPKETERTAHNLEIKYRCISINDRIQQLKVLCC-DEAKLSKATLRRAI 399
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 60 EYIQVRRKNHHTQDDIDDKR 81
QY 400 EHIEVEHENQVLKHHVQEMRK 421
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 11
ID W60569 standard; Protein; 346 AA.
AC W60569;
DE Human chromosome 19 derived USF2 gene product sequence.
KW Hydronephrosis gene; HNG gene; USF2 gene; renal disease; renal aplasia;
KW vesical-ureteral reflux; pelvi-ureteral junction obstruction;
KW multicystic renal dysplasia; renal agenesis; hydronephrosis;
KW Von Mayer-Rokitansky-Kuester disorder; bifid ureter.
OS Homo sapiens.
PN W09815650-A2.
PD 16-APR-1998.
PF 09-OCT-1997; E05583.
PA (VLAA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
PI Frys JPGJ, Groenen PMA, Van De Ven WJM;
DR WPI: 98-240833/21.
DR N-PSDB: V18130.
PT Hydronephrosis gene - useful to treat or diagnose renal diseases and
PT disorders, e.g. vesical-ureteral reflux, pelvi-ureteral junction
PT obstruction, multicystic renal dysplasia or renal agenesis
PS Disclosure; Fig 6A-B; 73pp; English.
CC This is a USF2 gene product derived from human chromosome 19 USF2 gene. A
CC translocation partner to this gene on chromosome 6 is the hydronephrosis
CC gene (HNG) product. The HNG gene can be used as a starting point to
CC design suitable compounds or techniques for the treatment of renal
CC diseases or disorders, or nucleotide probes for diagnosing cells involved
CC in renal diseases or disorders. A protein or a fragment encoded by HNG
CC gene can be used as a starting point for preparing suitable antibodies
CC

Query Match 1.5%; Score 118; DB 1; Length 352;
Best Local Similarity 28.3%; Pred. No. 6.47e+00;
Matches 34; Conservative 22; Mismatches 61; Indels 3; Gaps 3;

Db 33 DLEKKEGELMKEDDEDSLNHHNGENEEDEGDEDEDEDEDDDDDDKPKRRGPKKKM 92
QY 285 NMSPSEVERLNKKSGACSATNGPSRKEAKIVIQETAEDEDEDEDEDSGETMSQG 344
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```

for diagnosing cells involved in renal diseases and disorders. The products and method can be used to treat or diagnose renal diseases and disorders selected from vesical-ureteral reflux, uni or bilateral pelvi-ureteral junction obstruction, multicystic renal dysplasia, renal agenesis, renal aplasia, hydronephrosis, Von Mayer-Rokitansky-Kuester disorder and bifid ureter.

Sequence 346 AA;

Query Match 1.5%; Score 121; DB 1; Length 346; Best Local Similarity 34.9%; Pred. No. 4.05e+00; Matches 30; Conservative 19; Mismatches 29; Indels 8; Gaps 6;

Db 229 RTPDERRRAGHNEVERRRRKINNWIQVLSKIPDCNADNSLTGASGGILSKACDYIR 288
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 350 RPKTERTA-HNLEIKYRCISINDRIQQL-KVLL-CGDEAKL---SKSATLRRRAIEHIE 403
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 289 ELRQTNQRMQETFEAEER-LQMDNEL 313
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 404 EVEHENQVLKHHVQEMRKTLQ-NNRL 428

RESULT 12

ID R77505 standard; Protein; 352 AA.

AC R77505;

DE Frog neurogenic differentiation (NeuroD) protein.

KW NeuroD; neurogenic differentiation; neuronal growth factor;

KW basic helix-loop-helix secondary structure;

KW non-neuronal cell differentiation; antigen; drug screening;

KW neurodegenerative disease; traumatic injury; gene therapy.

OS Xenopus laevis.

EH Key Location/Qualifiers

FT domain 56..79

FT domain /note= "acidic domain"

FT domain 118..157

FT region /note= "HLH domain"

FT region 157..199

FT region /note= "highly conserved region"

PN W09530693-A1.

PD 16-NOV-1995.

PF 08-MAY-1995; U05741.

PR 06-MAY-1994; US-239228.

PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.

PA (WEIN/) WEINTRAUB N.

PI Hollenberg SM, Lee JE, Tapscott SJ, Weintraub HM;

DR WPI: 95-404081/31.

DR N-PSDB: T05514.

PT Nucleic acid molecule which hybridises with a neuroD HLH domain - is used in a method for inducing differentiation of a non-neuronal cell.

PS Claim 9; Page 39-40; 50pp; English.

CC NeuroD induces differentiation of a non-neuronal cell into a neuron. It is a member of a novel protein family and is transiently expressed in differentiating neurons during embryogenesis. DNA encoding NeuroD may be used in the development of probes, in the construction of recombinant cell lines and transgenic animals, and in the construction of gene therapy vectors for the repair of neuronal defects resulting from traumatic injury and neurodegenerative diseases (Alzheimer's disease, Huntington's disease, Parkinson's disease). Transformed host cells are used (1) as a source of neuronal growth factors, (2) in transient and continuous cultures for anti-cancer drug screening, and (3) as sources of recombinant NeuroD for use as an antigen in diagnostic antibody production.

CC Sequence 352 AA;

Query Match 1.5%; Score 118; DB 1; Length 352; Best Local Similarity 28.3%; Pred. No. 6.47e+00; Matches 34; Conservative 22; Mismatches 61; Indels 3; Gaps 3;

Db 33 DLEKKEGELMKEDDEDSLNHHNGENEEDEGDEDEDEDEDDDDDDKPKRRGPKKKM 92
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 285 NMSPSEVERLNKKSGACSATNGPSRKEAKIVIQETAEDEDEDEDEDSGETMSQG 344
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 93 TKARVERFKV-RRKANARENRMHGLNDALDSLRKVVPCYSKYTKLSTKETURLAKNYI 151
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 345 TTIIVRPKTERTTAHNLEIKKYRCSINDRIQOL-KVLCC-GDEAKLSKSATLRRRAIEHI 402

RESULT 13

ID W22437 standard; Protein; 352 AA.
AC W22437;
DE DT Frog neurogenic differentiation protein (NeuroD1).
DE Neurogenic differentiation protein; NeuroD; neuroD1 gene;
KW transcriptionsal activator; neuron; pancreas; gastrointestinal;
KW knock-out mouse; transgenic animal; cancer; diabetes; gene therapy.
OS Xenopus laevis.
FH Key Location/Qualifiers
FT FT 118..157 /label= HLH
FT FT /note= "helix-loop-helix region"
FT FT 157..199 /note= "highly conserved region characteristic
FT FT of NeuroD proteins"

PN WO9716548-A1.
PN 09-MAY-1997.
PF PD P30-OCT-1996; U17532
PR PR 02-NOV-1995; US-552142.
PA PA (HEIN-) HUTCHINSON CANCER RES CENT FRED.
PE PE (WEIN/) WEINTAUB N.
PI PI Hollenberg SM, Lee JE, Tapscott SJ, Weintraub HM;
PL PL WP1: 97-272117/24.
DR DR N-PSSB: T74888.

PT PT Nucleic acid encoding neurogenic differentiation polypeptide -
PT useful e.g. in regulating neuronal, endocrine and gastrointestinal
PT development
PS PS Claim 9; Page 57-58; 8lpp; English.

CC CC Human, mouse and Xenopus neurogenic differentiation proteins
CC CC (W22436-43) are transcriptional activators involved in neuronal,
CC CC endocrine and gastrointestinal development. They are active very
CC CC early in embryo development and may represent an/off switches,
CC CC possibly reversible, for controlling cell development. They were
CC CC discovered by expression cloning and screening assays designed to
CC CC identify basic-helix-loop-helix proteins capable of interacting
CC CC with the protein product of the Drosophila daughterless gene. Human,
CC CC mouse and frog neurod gene sequences (T74887-94) have been
CC CC identified. Cells expressing NeuroD polypeptides can be used to
CC CC screen for anticancer agents able to drive terminal differentiation
CC CC in neural tumours, to screen for (antagonists), or transported to a
CC CC site of neuronal injury or disease, and possibly (since NeuroD can
CC CC regulate insulin expression) to treat diabetics.

SQ SQ Sequence 352 AA;

•

Query Match 1.5%; Score 117; DB 1; Length 352;
Best Local Similarity 28.3%; Pred. No. 7.55e+00;
Matches 34; Conservative 22; Mismatches 61; Indels 3; Gaps

Db 33 DLEKEGELMKEDDESLNHHNGNEENEDEGGDEEDEDDEDDEDOKPFRGPKKXKM 92
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 285 NMSPSEVRLLKNKGSCATSINGPSRSKAIAKIQTAEAGDEDEDDSDSGETMSQG 344
: : | : | : | : | : | : | : | : | : | : | : | : | : | :

Db 93 TKARVERFKV-RRKANARENRMHGLNDALDSLRKVVPCYSKYTKLSTKETURLAKNYI 151
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |:
QY 345 TTIIVRPKTERTTAHNLEIKKYRCSINDRIQOL-KVLCC-GDEAKLSKSATLRRRAIEHI 402

RESULT 14

ID WT1010 standard; Protein; 352 AA.
AC WT1010;
DE DT 21-Oct-1998 (first entry)
DE Xenopus neuroD1 protein, which is a bHLH protein.
KW Basic helix-loop-helix; bHLH; neuroD; neuroectodermal tumour;
KW classification; medulloblastoma; frog.
OS Xenopus laevis.
FN FN US5795723-A.
PD PD 18-AUG-1998.

```
Best Local Similarity 23.8%; Pred.No.2.53e+00;
Matches 30; Conservative 25; Mismatches 67; Indels 4; Gaps 4;

Db      42 SHHPQHNNHHNNHHPPAPQPPPPQQOQQPAPQPPOARGAPADDDKGPQLLL 101
       |:| :| | | | | | | | | | | | | | | | | | | | | | | | | |
QY      221 SQHQSPPRRHHNNHHMPXIHENPEQVSPSIEDAETPKTHLVEQSPQNKEULL 280

Db      102 PPSAALDGAKADALGAKPGGGPAELAVGPDEKEKGAGAGEKKGAGEGKGEGGK 161
       |:| :| | | | | | | | | | | | | | | | | | | | | | | | | |
QY      281 RLLVNNSPSEVERLNK-KSGACSATNGP-SRS-KEKAARIVIQET-ABGDDEDEDSD 336

Db      162 EGDKN 167
       |:| :| :|
QY      337 SGETMS 342
```

Search completed: Sun Sep 3 12:33:46 2000
Job time : 48 secs.

Query Match	5.3%	Score 421	DB 2	Length 1141
Best Local Similarity	25.1%	Pred. No. 4,4/e-51		
Matches 190	Conservative 210	Mismatches 280	Indels 77	Gaps 59
Db	326	PK-TERRAHMLIEKTRKCSINDRIQOLKYLICGDEAKLSKATLRAIEHT--EEVEH	407	
Db	386	KLRQENNVYLKLANOKNKLKIDIGSLVDNEVDLIKIDFENONVYLMSPADSGSQAFS	445	
Qy	408	----ENQYLK--HHVEQMRKTLQNNRLPYRPE-PIQYETYSARSPVESPPSPRNERK--R	458	
Db	446	PYSIDSEGPSLLDDAKYKDEPDSPYALGKVDNRSLILCVLFLCSFNPLTSLLOWGG	505	
Qy	459	S-RM-S-TTTPM-RNGT-RD--GSSKYLTL-F--A-MLLAVLFPNPGLLAGSAIFSKAA	506	
Db	506	AHDSQHHSSSGSSVLSFESGSG-GWF-D-WMMPTLLMLVNOVYILSVFYKLAVGER	562	
Qy	507	AEAPLASFE-HGRYIDDPDGTSTRLLFWESIIIMSYWFWNLMTIYVVKLLIHGDP	565	
Db	563	VIRPHRSRSYTFEMHRKQADLDLRGDFAAAAANILQTLCLAVLGRPLTSRLDLACLSMN	622	
Qy	566	VQDFMSWMOQFVYTTREARARLNSGNLKDQAKRFCECLATLDLRSLPSPGVDSVFSVGWE	625	
Db	623	VIRSLQKRLVRLWMLKKVFOCRBRATPATEGAFEDAKTSARDAALAYHRHLQHL--IT	679	
Qy	626	CVRH-LLN-WL--WI-GR-YIARRRRSTRKPV-SVQCRSHQOTAVL-YHEIHQHLMGIT	677	
Db	680	GLPL-A-GSAC-SQVHMLCAVNLAECA--EEKIPPSSTVEIHLTAMGLKTRCGSKLG	733	

[illegible]

```

#cross-references MUID:94274723
#accession A54164
#status preliminary
#molecule-type mRNA
#residues 1-1133 ##label SAT
#cross-references GB:U09103; NID:g516002; PIDN:AAA20085.1; PID:g516003
CLASSIFICATION #superfamily sterol regulatory element binding protein
SUMMARY #length 1133 #molecular-weight 120464 #checksum 4586

Query Match 4.8%; Score 387; DB 2; Length 1133;
Best Local Similarity 29.4%; Pred. No. 7,856-45;
Matches 149; Conservative 132; Mismatches 171; Indels 54; Gaps 41;

Db 313 OSREKPTAHAAIEKRYRSSINDKIVELKDLVGTGTEAKLNKSAVLRAKIDYRFLQHSN 372
   :::::::::::::::::::::
Qy 351 RPKTERRAHNLIEKKYRCSINDRIQOLKVLGDEAKLSATIRAEIHEVEHENQ 410
   :::::::::::::::::::::
Db 373 KLGQENALRNAHKSLSKDLVSACGSAGGTDAVEGVKPEVDTLTPPSDAGSPSOS 432
   :::::::::::::::::::::
Qy 411 VLKHHVEQMRKTLQNNR-L-PY-PE--PIQYTESARS-P--VES-SPSP-RNERKRS 459
   :::::::::::::::::::::
Db 433 SPLSGSGSSGSDSPDPSPFEDSOVKQRLHSHGMDLRSLALCALVFLCITCNPL 492
   :::::::::::::::::::::
Qy 460 R-MSTTTP-MKNGTRDSSKVTLFA-MLLAVLIENPGLAGSAT-F-SKA-A-AE-API 511
   :::::::::::::::::::::
Db 493 ASLFGWGIPIGPSASAGHSSGRSMLAEESRDGSMWTOMLPLPLVWLANGLVLACIAL 552
   :::::::::::::::::::::
Qy 512 ASPEHG-RVIDDDPGT--ST-RTLF---WEGSI-IN--MS-YVWFNIMLIIVYVKL 559
   :::::::::::::::::::::
Db 553 FVYGEVPTRPSPAVH-FMRHRKQADLDLARGDFAQAQOLWALQALGRPLTNSDL 611
   :::::::::::::::::::::
Qy 560 LIHDPV-QDMSVSWQTFVTTRKARAEINSGNLKDAQRFCECLATLDRLSPGVDS 618
   :::::::::::::::::::::
Db 612 ACSLLMNIIRHLQRLMWGRMLAGAGLRDCCGLMRASARPAALYHKLHGLHMG 671
   :::::::::::::::::::::
Qy 619 VFSVGMCEVRLHLMWMLGRIYARRRSTTRKPVSV-V-CRSHAQTAVL-YHEIHQLHMG 675
   :::::::::::::::::::::
Db 672 --KY--T-G-GHLIASNLASLNLAECAG--DAVSMATLAETIYVAALRKVSLPRA 721
   :::::::::::::::::::::
Qy 676 ITGNFEDIEFSALTGLEMSICAVNLAAGASNDGLPRAVMAQIYISASIOCRALPNL 735
   :::::::::::::::::::::
Db 722 L-HFLTRFELSSAR-QACLAOSGAVPLAMOWLCPVGRFFVDDMAVHG-AP-QESL-Y 776
   :::::::::::::::::::::
Qy 736 LAPFSGCYFLRRARRHVRAREHSVS-HLMTFHPATRKFMSDAKRLHEVLSSKQRLRF 794
   :::::::::::::::::::::
Db 777 -S-VAGNPVDPLAOVTRFLCEHLER 800
   :::::::::::::::::::::
Qy 795 GSFVEDEQLSPLARIRITLKYLLSK 820
   :::::::::::::::::::::

RESULT 4
ENTRY A48085 #type complete
TITLE transcritpion factor Addl - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 26-May-1994 #sequence_revision 26-May-1994 #text_change 1-Sep-1999
ACCESSION A48085
REFERENCE A48085
#authors Tontozou, P.; Kim, J.B.; Graves, R.A.; Spiegelman, B.M.
#journal Mol. Cell. Biol. (1993) 13:4753-4759
#title Addl: a novel helix-loop-helix transcription factor
associated with adipocyte determination and
differentiation
#cross-references MUID:93330269
#accession A48085
#status preliminary
#molecule-type mRNA
#residues 1-927 ##label TON
#cross-references GB:L16995
CLASSIFICATION #superfamily sterol regulatory element binding protein
KEYWORDS #leucine zipper
SUMMARY #length 927 #molecular-weight 98517 #checksum 1870

```

```

Query Match 4.7%; Score 375; DB 2; Length 927;
Best Local Similarity 30.2%; Pred. No. 1,206-42;
Matches 155; Conservative 125; Mismatches 180; Indels 53; Gaps 37;

Db 282 GKALGASQSGKEKRTAHNAIEKRYRSSINDKIVELKDLVGTGTEAKLNKSAVLRAKIDYR 341
   :::::::::::::::::::::
Qy 344 GTTILVRPKTERRAHNLIEKKYRCSINDRIQOLKVLGDEAKLSATIRAEIHE 403
   :::::::::::::::::::::
Db 342 FLQHSNOKLQENALRNAHKSLSKDLVSACGSAGGTDAVEGVKPEVDTLTPPSDA 401
   :::::::::::::::::::::
Qy 404 EEVHENQVLAHVEQMRKTLQNN--R-LPY-EPLOYTEYS--ARSP--VES-SPSP-R 453
   :::::::::::::::::::::
Db 402 GSPSGSPLSGSGSSGSDSPDPSPFEDNOVKAQRLSHSGMDLPSRLALCVLF 461
   :::::::::::::::::::::
Qy 464 NERKRS-MSTTTP-MKNGTRDSSKVTLFA-MLL-AV-LFPNPLGLAGS--A--IRS- 503
   :::::::::::::::::::::
Db 462 LCLTNPPLASLFWGILTPSDASGVHSSGRSMLAEESRDGSMWTOMLPLPLVWLANGL 521
   :::::::::::::::::::::
Qy 504 KAAAPAPIASPFEHGRV-L-DDPDGT--ST-RTLF---WEGSI-IN--MS-YVWFNIM 551
   :::::::::::::::::::::
Db 522 VLACALLFYGEVPTRHSGPAVHFMHRKQADLDLARGDFAQAQOLWALQALGRPL 581
   :::::::::::::::::::::
Qy 552 IIVYVVKLLIHGDVQDPMVSWQTFVTTRKARAEINSGNLKDAQRFCECLATLDRL 611
   :::::::::::::::::::::
Db 582 PTSNLDLACSLMNLVRLHQLRWGRMLAGAGLQDRYRLRDARASARDAAVYHKL 641
   :::::::::::::::::::::
Qy 612 PSPGVDSVFSVGMCEVRLHLMWMLGRIYARRRSTTR--PVSVCRRHAQ-TAVLYHEI 668
   :::::::::::::::::::::
Db 642 HOLHMG---KYTGGH-LVA-SNL--ALSALNLAECAG--DAISMATLAETIYVAALRV 691
   :::::::::::::::::::::
Qy 669 HOLHMGITGNFEDIEFSALTGLEMSICAVNLAAGASNDGLPRAVMAQIYISASIOQ 728
   :::::::::::::::::::::
Db 692 KTSILPRAL-HFLTRFELSSAR-QACLAOSGAVPLAMOWLCPVGRFFVDDMAVHG-AP 748
   :::::::::::::::::::::
Qy 729 RLALPNLLAPFSGCYFLRRARRHVRAREHSVS-HLMTFHPATRKFMSDAKRLHEVLSS 787
   :::::::::::::::::::::
Db 749 -QESL-Y-S-VAGNPVDPLAOVTRFLCEHLER 777
   :::::::::::::::::::::
Qy 788 KQQLRFGSFEDEQLSPLARIRITLKYLLSK 820
   :::::::::::::::::::::

RESULT 5
ENTRY T13348 #type complete
TITLE transcritpion factor H1H106 - fruit fly (Drosophila
melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 20-Sep-1999
ACCESSION T13348
REFERENCE T13348
#authors Theopold, U.; Ekengren, S.; Hultmark, D.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:1195-1199
#title H1H106, a Drosophila transcription factor with similarity to
the vertebrate sterol responsive element binding protein.
#cross-references MUID:9616553
#accession T13348
#status preliminary; translated from GB/EMBL/DBJ
#molecule-type mRNA
#residues 1-1113 ##label THE
#cross-references EMBL:U38238; NID:g1079655; PID:g1079656;
PIDN:AAA97864.1
CLASSIFICATION #superfamily sterol regulatory element binding protein
SUMMARY #length 1113 #molecular-weight 14513 #checksum 539

Query Match 4.0%; Score 322; DB 2; Length 1113;
Best Local Similarity 24.6%; Pred. No. 3,806-33;
Matches 192; Conservative 204; Mismatches 305; Indels 81; Gaps 68;

Db 261 SPANFVOGKYPINRQPKVEKRSANHAIEKRYRSSINDKINELKNLVVGEQAKLNKS 320
   :::::::::::::::::::::
Qy 337 SGETWS-OGTTIIVR-RPKT-E-RTAHNLIEKKYRCSINDRIQOLKVLGDEAKLSK 392
   :::::::::::::::::::::
Db 321 AVLKRSIDKINDIQRQNDKAKELQRLQRELMDAGSVKXKLLDLGTRPGASKKRRSS 380
   :::::::::::::::::::::

```

QY	393	ATIRALAIHIEVEHENQVTK--HHVE-Q-M-RK-T-LQNN-RLPY-P-P-EPIYTEFS	440
Db	381	QTFTTDACTIPRDESDPGLSPHMSDLSLPPSYGGSTACSSSSSSNEEPLVPPSM	440
QY	441	ARSPVSSPSR-NEKRS--RM-STTT-PMKN-GTRDGS-SKTYLPAMLAVLIFNFI	493
Db	441	RGMAHSRLGLCMEMFALLAVNPFKTLFQRGYSDNDLGDMSGGORILISYVEEGFAY	500
QY	494	-GLLA-GS-AL--FSKAA-AEAPLISPEHGVYIDPD-GT-ST--RTL-FW-EG-SI-I	539
Db	501	WOOSWIMLNFNTLMAGCLVLYLGVDP-Q-LDAQT-DAYCOHROADFEFSOGSSQAY	557
QY	540	-NMS-YVWVFNTIMIIYVVVYKLLIHGDPPQDMYSVSWQTFVTRKARALNSGMLKDAQ	597
Db	558	AGYLNCIMFGSLSPASRECIYLOTTWQFLRFLRLMIGRLVSRSGLSFNAASKRQA	617
QY	598	RKFCCCLATLDRSLSPGVDYFVSGWECVRLHMLMWLGRIARRRRTTKPVSGNS	657
Db	618	LASARELALFENRLQLOLTGN--GS--RGDNGIMMALFASMAEVA--HNLTLPRETI	671
QY	658	HAQTVLYHEIHOQLHMGITGTNGFEDTIEPSALTGLFMSLCANVLAEMAGASMDGILPRAV	717
Db	672	C-IHYMTALRMKRSAPKWLQDFPARYMSRAHOCGRTRATBOT-OELMAETAYGYR-	728
QY	718	AOIYISASIOCLRALPNLPAFFSGYFLRRARRH-R-RAPEHSVHLLIMIPATRKFM	775
Db	729	CATHTFTYDLSGSGODGFTFLRN-PCDPAHVAKQYREHLLFSIOCLVAGAKHSGGL	787
QY	776	SDAKRLEHVLSSKQKOLRFGSFVEDEQLSPILARITTKVLLSKLYOELV-GDEIFTK	834
Db	788	PTSSVSGAEQLOQO-Q--HSGTIVSNV-LKYSLL-KDT-LMAD-ED--ERDTNV--VM	836
QY	835	NVERLINDNRDDEVDVVDVSRLLVTISOCALITNKDSAFEGYISNSGACCMW	894
Db	837	MADVLETAHVMLIGEDTLAEOLYGRIKQPTLOQCGENDHLPKALHAYLRAMKMLTKNN	896
QY	895	WTHVLTGCIYGRSNKNEELARQHSYLSIRNCPKIL--TDNLGLAVG-HA-LCARAKICIDR	950
Db	897	GNALDKSLKQLYNICDESSVLEOECTVYNRITDA-KGIKLLEFOLLTCWMLLETATLME	955
QY	951	-DS-PK-VSOYVCYHTKKSLESLELRFTSSRSAGVSGIOESTRRMAEWINSLIDAR	1007
Db	956	LEHNMEDGFF-QY-PGE-VLEK-FQTDLNLARNIVENIPASORITLYEAVCRILMGA	1011
QY	1008	-SNL-FASKPYWTQSGFKGSGTSTLYQEAANYN-ALINTRODCMRLLFYELTCMLNGA	1064
Db	1012	SP 1013	
QY	1065	NP 1066	
RESULT	6		
ENTRY	PD0035	#type complete	
TITLE		sterol regulatory element-binding protein-1 isoform - mouse	
ORGANISM		#formal_name Mus musculus #common_name house mouse	
DATE	23-Jul-1999	#sequence_revision 23-Jul-1999	#text_change 20-Aug-1999
ACCESSIONS	PD0035		
REFERENCE	PD0035		
#authors	Inoue, J.; Sato, R.		
#journal	Biosci. Biotechnol. Biochem. (1999) 63:243-245		
#title	A novel splicing isoform of mouse sterol regulatory element-binding protein-1 (SRBP-1).		
#cross-references	MUID:99161303		
#accession	PD0035		
#status	preliminary		
#molecule type	mRNA		
#residues	1-391	#label	INO
#cross-references	DDbJ:AB017337; NID:g4240011		
KEYWORDS	transcription		
SUMMARY	length 391 #molecular_weight 39912 #checksum 4894		

	Query Match	3.3%	Score 265;	DB 2;	length 391;
	Best Local Similarity	50.0%	Pred. No.	2.84e-23;	
	Matches	39;	Conservative	18;	Mismatches 21; Indels 0; Gaps 0;
Db	268	OSRGEKPTAHNAIEKPYRSSINDKIVEIKDLVGTGTEAKLNKSAYLRKAIDYIRFLQHSNQ	327		
Oy	351	RPKETRRRAHLLEIKTKTRCISNDINIQLVLCODEKSKLSATLRAAIEHIEVEHENQ	410		
Db	328	KLKQENTLTLSAAHSKSL	345		
Oy	411	VLKHHVEDQMRTKLNNRL	428		

RESULT	7
ENTRY	A40728
TITLE	#type complete microphthalmia-associated transcription factor mi - mouse
ORGANISM	#formal_name Mus musculus #common_name house mouse
DATE	19-May-1994 #sequence_revision 19-May-1994 #text_change

DATE _____

ACCESSIONS	A40728; A48021; I49245; S36645
REFERENCE	A40728

*Journal
#40001
Call (1903) 74-305-404
E.; Copeland, N.G.; Jenkins, N.A.; Arnheiter, H.
Houghlinson, C.A.; Moore, A.O.; Nakayama, A.; Steinglimsson
*auctois

*Journal
#title
Mutations at the mouse microphthalmia locus are associated with defects in a gene encoding a novel basic-helix-loop-helix-zipper protein.

```
#cross-references MUID:93345026
#accession      A40728
```

```
##molecule_type mRNA
##residues 1-419 ##label HOD
```

#CROSS-References GR: 223000; MID: g3390/39; PID: g3390/40
REFERENCE A48021

```
#journal J. Biol. Chem. (1993) 268:20687-20690
#title A helix-loop-helix transcription factor-like gene is located
```

#cross-references MUID:94012591
#accession A48021

```
##status      preliminary
##molecule_type DNA
```

##cross-references GB:L22958
REFERENCE I49244

AUTHORS
Steingrimsson, E., Moore, A.O., Lamotte, M.J.,
Ferrel-D'Amaré, A.R.; Burley, S.K.; Sanders Zimring, D.C.;
Skow, L.C.; Haddockson, C.A.; Arnharter, W.; Conneland

N.G.; Jenkins, N.A.
Nature Genet. (1994) 8:256-263

explain their developmental and phenotypic consequences.
#cross-references MUID:95179171

```

#accession 143243
##status preliminary; translated from GB/EMBL/DBJ
##refcount 4
#num span

```

```
#residues      1-16  #label RES
#cross-references EMBL:U19875; NID:g642572; PID:g642573
```

#gene
KEYWORDS
mi
DNA binding, transcription factor

SUMMARY #Length 419 #molecular-weight 46768 #checksum 1668

Best Local Similarity	31.28;	Pred. No. 3.81e-05;	
Matches	24: Conservative	27: Mismatches	21: Indels
			5: Gaps

Db 204 QKDNHNLIERRRRNINDRIELGTLIPKSNDDPMRWKGTILKASVDYIRKLQREQR 263

355 ERTAHNIEKKYRCSINDRIQOLVLLC-GDEAKL--SKSATLRAIEHIEVEHENOV 411

— :: — :: — :: — :: — :: —

OY 412 LKHVQMKRTIQN-NR 427

RESULT 8

ENTRY 138024 #type complete

TITLE MTF protein - human

ORGANISM #formal_name Homo sapiens #common_name man

DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change 29-Aug-1997

ACCESSIONS 138024

REFERENCE 138024

#authors Tachibana, M.; Perez-Jurado, L.A.; Nakayama, A.; Hodgkinson, C.A.; Li, X.; Schneider, M.; Mikl, T.; Fex, J.; Francke, U.; Arnheiter, H.

#journal Hum. Mol. Genet. (1994) 3:553-557

#title Cloning of MTF, the human homolog of the mouse microphthalmia gene and assignment to chromosome 3p14.1-p12.3

#cross-references MUID:94348499

#accession 138024

#status preliminary; translated from GB/EMBL/DBJ

#molecule_type mRNA

#residues 1-419 ##label RES

GENETICS

#cross-references EMBL:229578; NID:9468496; PID:9468497

#gene GDB:MTF

#cross-references GDB:214776; OMIM:156845; OMIM:193510

#map_position 3p14.1-3p12

SUMMARY #length 419 #molecular-weight 46938 #checksum 307

Query Match 1.9%; Score 149; DB 2; Length 419;
Best Local Similarity 31.2%; Pred. No. 3.81e-05;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

DB 204 OKDNHNIERRRRNINDRIKELGTLIPKSNPDMRNKGTILKASVYIKLQREQOR 263

OY 335 ERRAHNIIEKKYKRSINDRIQKLVLC-GDEAKL--SKSATLRRAIEHIEVEHENOV 411

DB 264 AKE-LENROKLEHNR 279

OY 412 LKHVQMKRTIQN-NR 427

RESULT 9

ENTRY 114752 #type complete

TITLE microphthalmia-associated transcription factor, MTF-A - human

ORGANISM #formal_name Homo sapiens #common_name man

DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Sep-1999

ACCESSIONS 114752 JE0207

REFERENCE 218180

#authors Koehner, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

#submission submitted to the Protein Sequence Database, August 1999

#accession 114752

#status preliminary

#molecule_type mRNA

#residues 1-520 ##label KOE

#cross-references EMBL:AL10195

#experimental_source adult uterus; clone DKFZP586B217 JE0207

REFERENCE

#authors Amai, S.; Fuse, N.; Yasumoto, K.; Sato, S.; Yajima, I.; Yamamoto, H.; Usono, T.K.; Durlu, Y.; Tamai, M.; Takahashi, K.; Shibahara, S.

#journal Biochem. Biophys. Res. Commun. (1998) 247:710-715

#title Identification of a novel isoform of microphthalmia-associated transcription factor that is enriched in retinal pigment epithelium.

#cross-references MUID:98321192

#accession JE0207

#molecule_type mRNA

#residues 1-520 ##label AMA

#cross-references DBJ:AB006909

COMMENT This protein plays a important role in differentiation of retinal pigment epithelium and regulates melanogenesis in retinal pigment epithelium.

GENETICS

#note DKFZ586B217.1

SUMMARY #length 520 #molecular-weight 58162 #checksum 6813

Query Match 1.9%; Score 149; DB 2; Length 520;
Best Local Similarity 31.2%; Pred. No. 3.81e-05;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

DB 305 OKDNHNIERRRRNINDRIKELGTLIPKSNPDMRNKGTILKASVYIKLQREQOR 364

OY 335 ERRAHNIIEKKYKRSINDRIQKLVLC-GDEAKL--SKSATLRRAIEHIEVEHENOV 411

DB 365 AKE-LENROKLEHNR 380

OY 412 LKHVQMKRTIQN-NR 427

RESULT 10

ENTRY B36289 #type complete

TITLE CeyYOD protein - Caenorhabditis elegans

ORGANISM #formal_name Caenorhabditis elegans

DATE 08-Mar-1991 #sequence_revision 08-Mar-1991 #text_change 23-Feb-1997

ACCESSIONS B36289

REFERENCE B36289

#authors Krause, M.; Fire, A.; Harrison, S.W.; Priess, J.; Weintraub, H.

#journal Cell (1990) 63:907-919

#title CeyYOD accumulation defines the body wall muscle cell fate during Caenorhabditis elegans embryogenesis.

#cross-references MUID:91077929

#accession B36289

#status preliminary

#molecule_type DNA

#residues 1-320 ##label KRA

#cross-references GB:M37497

#note the authors translated the codon TTG for residue 258 as Lys

KEYWORDS DNA binding; nucleus; transcription regulation

SUMMARY #length 320 #molecular-weight 36032 #checksum 2241

Query Match 1.8%; Score 141; DB 2; Length 320;
Best Local Similarity 25.6%; Pred. No. 4.61e-04;
Matches 32; Conservative 35; Mismatches 49; Indels 9; Gaps 7;

DB 110 EKSTPNATELIOSRVDSDHEDTTSTAGAGVG-PRRTKIDRRKAATMRERRLR-KV 167

OY 313 EKAATVIOETAEDEDEDESDSGETMSQSTIIIVRPKERTTAHLIRK-YRCSI 371

DB 168 NEAFVVKQRTCPNPQRLPKYELIRSAIDYINLNR--ML-QQAGKTKIMEQNOHLQ 223

OY 372 NDRIQKLVLCGD-EAKLSKATLRRAIEHIEVEHENOV LKHVQMKRTI-QNNRLP 429

DB 224 MTQOI 228

OY 430 YPEPI 434

RESULT 11

ENTRY A42029 #type complete

TITLE transcription factor E3 - mouse

ORGANISM #formal_name Mus musculus #common_name house mouse

DATE 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 29-Jan-1999

ACCESSIONS A42029

REFERENCE A42029

#authors Roman, C.; Matera, A.G.; Cooper, C.; Artandi, S.; Blain, S.; Ward, D.C.; Calame, K.

#journal Mol. Cell. Biol. (1992) 12:817-827

```

#title      mTFE3, an X-linked transcriptional activator containing basic helix-loop-helix and zipper domains, utilizes the zipper to stabilize both DNA binding and multimerization.
#cross-references MUID:92123207
#accession  A42029
##molecule_type mRNA
##status     Preliminary
##residues   1-446 ##label ROM
##cross-references GB:S76673; NID:g243439; PID:g243440
##note       sequence extracted from NCBI backbone (NCBIN:76673, NCBI:P76674)
#length 446 #molecular-weight 47891 #checksum 8873

SUMMARY
Query Match          1.8%; Score 143; DB 2; Length 446;
Best Local Similarity 31.2%; Pred. No. 2,49e-04;
Matches 24; Conservative 26; Mismatches 22; Indels 5; Gaps 4;

Db    219 OKKDNHLEERRRRFNNRIKELGFLIPKSNPEMKRMNKGTLLIKASVDYIRKLQKEQR 278
      ::: |||||: | :|||: | | :|::: ::||: |::| |::| |
Oy    355 ERTAHNLIEKKTKRCINDRIGQLKVLCC-GDEAKL--SKSALLRLRAIETHEVEHENOV 411
      | :|::| | | |
      412 LKHNVQMRKTL-QNNR 427

RESULT 12
ENTRY      A54743 #type complete
TITLE      transcritpion factor HFK1 - human
ORGANISM   Homo sapiens #common_name man
DATE       18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 12-Sep-1997

ACCESSIONS A54743
REFERENCE   A54743
AUTHORS     Murphy, D.B.; Wiese, S.; Burfeind, P.; Schmundt, D.; Mattei, M.G.; Schulz-Schaeffer, W.; Thies, U.
JOURNAL     Genomics (1994) 21:551-557
TITLES      Human brain factor 1, a new member of the fork head gene family.

#cross-references MUID:95048332
#accession  A54743
#status     preliminary; nucleic acid sequence not shown
##molecule_type RNA
##residues  1-476 ##label MUR
##cross-references GB:X74142
GENETICS    GDB:FKNL4; HBF-1; HFK1
#gene       ##cross-references GDB:433550
#map_position 1q12-14q12
CLASSIFICATION superfamily unassigned fork head proteins; fork head DNA-binding domain homology

FEATURE
169-260     #domain fork head DNA-binding domain homology #label FHD
SUMMARY     #length 476 #molecular-weight 50830 #checksum 5990

Query Match          1.8%; Score 145; DB 2; Length 476;
Best Local Similarity 29.1%; Pred. No. 1.34e-04;
Matches 37; Conservative 27; Mismatches 55; Indels 8; Gaps 8;

Db    39 HNNSHNPNNNNNNNNNRRRARPRRRRAQQORRRPRLAROGAQSNDKGQP 98
      ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Oy    223 HNOSPPRNHHNRRMRKINENRPVAVSPSTEDARETRTHL-VERPSPKSPNNKEELLR 281
      ||| | | | | | | | | | | | | | | | | | | | | | | | | |
      99 LHL-LPTDHNRRPSGAKAGGCSRPGLGVGFDEKKGAGAGEEGKGAEGSGDEGG 157
      ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Oy    282 LLVNMSPEVERL-KKKKSACS-ATN-GP-SNS-KRKAKIVQT-ABEDDEDDEDS 335
      | | | |
      158 KEGEKKN 164
      | | | |
Oy    336 DSGETWS 342

RESULT 13
```

```

ENTRY      A34596      #type fragment
TITLE      transcritpion factor E3 - human (fragment)
ORGANISM   Homo sapiens #common_name man
DATE       06-Jul-1990 #sequence_revision 09-Oct-1992 #text_change
                                     24-Sep-1998

ACCESSIONS
REFERENCE   A34596; S10379
#authors   Beckmann, H.; Su, L.K.; Kadesch, T.
#journal   Genes Dev. (1990) 4:167-179
#title     TF3: a helix-loop-helix protein that activates transcription
            through the immunoglobulin enhancer muE3 motif.
#cross-references MUID:90249724
#accession A34596
##status   preliminary
##molecule_type mRNA
##residues 1-536 ##label BEC
##cross-references EMBL:X5130; NID:g37061; PID:e20907; PID:g135343
#note      the authors translated the codon ACC for residue 433 as
            Ser, and GAG for residue 472 as Gln

GENETICS
#gene      GDB:TFE3
#cross-references GDB:125670; OMIM:314310
#map_position Xp11.23-Xp11.22
KEYWORDS   DNA binding; transcription factor
SUMMARY    #length 536 #checksum 8782

Query Match      1.8%; Score 143; DB 2; Length 536;
Best Local Similarity 33.8%; Pred. No.2,49e-04;
Matches 26; Conservative 23; Mismatches 23; Indels 5; Gaps 4;

Db 139 OKKNHNLIRRRRPINIDRIKELGTLIPKSSPDEKRWNKGTILKASVDYIRKLQEQOR 198
      :::|||||:::|||||:::|:::|:::|:::|:::|:::|:::|
Oy 335 ERRAHMLIEKKTRCSINDRIQLKVL--CGD-EAKLSKSTLRALIEHIEVHEHNOY 411
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 199 SKD-LESQRSLEQANR 214
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 412 LKHVEQMRKTL-QNNR 427

RESULT 14
ENTRY    S08343      #type complete
TITLE    nodulin precursor - soybean
          #formal_name Glycine max #common_name soybean
          29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change
DATE     20-Aug-1999

ACCESSIONS
REFERENCE   S08343; S08344; A27059
#authors   Franssen, H.J.; Thompson, D.V.; Idler, K.; Kormelink, R.; van
            Kammen, A.; Bisseling, T.
#journal   Plant Mol. Biol. (1989) 14:103-106
#title     Nucleotide sequence of two soybean ENOD2 early nodulin genes
            encoding Ngm-75.
#cross-references MUID:91332483
#accession S08343
##molecule_type DNA
##residues 1-309 ##label FRA
##cross-references EMBL:X16876; NID:g18575; PIDN:CAA34758.1; PID:g18576
#experimental_source strain Wayne
#genetics   CH1
#accession  S08344
##molecule_type DNA
##residues 1-309 ##label FRZ
##cross-references EMBL:X16876; NID:g18578; PIDN:CAA34759.1; PID:g18579
#experimental_source strain Wayne
#genetics   CH2
REFERENCE   A94164
#authors   Franssen, H.J.; Nap, J.P.; Gloudeemans, T.; Striekema, W.; van
            Dam, H.; Govers, F.; Louwerse, J.; van Kammen, A.;
            Bisseling, T.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1987) 84:4495-4499
#title     Characterization of cDNA for nodulin-75 of soybean: a gene
            product involved in early stages of root nodule
            development.

```


THIS PAGE BLANK (USPTO)

 NWSEIET (TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MsrchLpp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:28:01 2000; Maspar time 80.17 Seconds

Tabular output not generated. 962.525 Million cell updates/sec

Title: >US-09-332-522B-2
 Description: (1-1113) from US09332522B.pep
 Perfect Score: 8012
 Sequence: 1 MNEEFEGDVPMSDPFLSLVT.....AOPDAFHLLTVKLTSMDL 1113

Scoring table: PAM 150
 Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database:

sptrembl12
 1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
 5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
 9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
 13:sp_vertebrate 14:sp_virus

Statistics: Mean 54.388; Variance 102.303; scale 0.532

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	7998	99.8	1113	5	Q9XX00 Y4703B.7 PROTEIN.	0.00e+00
2	322	4.0	1113	5	Q24146 HLH106.	1.18e-36
3	265	3.3	403	11	Q9WNT3 STEROL REGULATORY ELEM	7.73e-26
4	229	2.9	144	6	Q97676 ADIPOCYTE DETERMINATIO	2.80e-19
5	161	2.0	498	5	Q91527 SIMILARITY TO A HELIX-	8.56e-08
6	149	1.9	110	11	Q88368 MICROPHTHALMIA ASSOCIA	6.03e-06
7	149	1.9	377	11	Q70241 MITF (FRAGMENT).	6.03e-06
8	149	1.9	419	4	Q14841 MITF PROTEIN (MICROPH	6.03e-06
9	149	1.9	468	13	Q73871 MITF.	6.03e-06
10	149	1.9	474	11	Q08874 MICROPHTHALMIA-ASSOCIA	6.03e-06
11	149	1.9	520	4	Q75030 A-TYPE MICROPHTHALMIA	6.03e-06
12	151	1.9	1829	11	Q35889 L-AFADIN.	3.00e-06
13	139	1.7	281	5	Q21663 SIMILAR TO HUMAN MAX I	1.83e-04
14	135	1.7	317	11	Q9WNT4 TRANSCRIPTION FACTOR T	6.92e-04
15	140	1.7	320	10	Q9ZUV8 PUTATIVE PLASOLIN G-B	1.31e-04
16	136	1.7	347	4	Q14948 TREC ISOFORM (OR TREC	4.98e-04
17	137	1.7	567	5	Q15817 PAXILLIN-LIKE PROTEIN.	3.57e-04
18	125	1.6	223	5	Q9XZ11 MYOD HOMOLOG (FRAGMENT	1.73e-02
19	126	1.6	310	11	Q35410 TRANSCRIPTION FACTOR U	1.26e-02
20	126	1.6	310	11	Q09135 USF1 PROTEIN.	1.26e-02

21	127	1.6	338	4	Q60409 AP-4.	9.19e-03
22	129	1.6	380	5	Q94749 SERINE-RICH PROTEIN (F	4.86e-03
23	130	1.6	391	10	P93667 BL2-1 PROTEIN.	3.53e-03
24	131	1.6	405	10	Q03462 OPAQUE2 HETERODIMERIZI	2.55e-03
25	131	1.6	410	10	Q41786 OPAQUE2 HETERODIMERIZI	2.55e-03
26	126	1.6	415	10	Q23192 HYDROTICAL 45.2 KD P	1.26e-02
27	127	1.6	417	5	Q02080 SIMILAR TO CCAAT/ENHAN	9.19e-03
28	132	1.6	434	3	Q12398 CHROMOSOME XV READING	1.85e-03
29	126	1.6	524	10	Q08036 F14J9.19 PROTEIN.	4.86e-03
30	129	1.6	604	13	Q42472 SYNDACAN-1 (FRAGMENT).	1.26e-02
31	126	1.6	675	5	Q61310 TSJ5.	1.26e-02
32	131	1.6	1212	5	Q24523 BUNCHED PROTEIN, CLASS	2.55e-02
33	132	1.6	1816	4	Q75088 AF-6.	1.85e-03
34	124	1.5	219	13	Q91151 MYOGENIC REGULATORY FA	2.36e-02
35	124	1.5	405	10	Q41757 OPAQUE-2 HETERODIMERIZ	2.36e-02
36	121	1.5	509	13	Q91236 ZONA PELLUCIDA PROTEIN	5.97e-02
37	123	1.5	592	5	Q9XV50 F29C12.1 PROTEIN.	3.22e-02
38	124	1.5	790	10	Q23068 SIMILAR TO RECEPTOR-LI	2.36e-02
39	121	1.5	876	2	Q45458 DNA POLYMERASE I.	5.97e-02
40	122	1.5	974	10	Q49634 HYPOTHETICAL 109.0 KD	4.39e-02
41	120	1.5	1004	2	P71719 HYPOTHETICAL 105.6 KD	8.11e-02
42	124	1.5	1267	5	Q93564 F45H11.4 PROTEIN.	2.36e-02
43	123	1.5	1271	5	Q21789 R07B5.8 PROTEIN.	3.22e-02
44	124	1.5	1743	4	Q75087 AF-6.	2.36e-02
45	123	1.5	3164	14	Q69088 VIRION PROTEIN.	3.22e-02

ALIGNMENTS

RESULT 1
 ID Q9XX00 PRELIMINARY; PRT; 1113 AA.

AC Q9XX00;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)

DE Y47D3B.7 PROTEIN.

GN Y47D3B.7.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditicoidea; Rhabditidae; Peloderinae; Caenorhabditis.

```

Db 61 SPPOEYDIDGORDVSTLHSLNHNNDFFSMRFSPPNFDLGGGRPSLAATQOLSGEGP 120
   |||||||
Qy 61 SPPOEYDIDGORDVSTLHSLNHNNDFFSMRFSPPNFDLGGGRPSLAATQOLSGEGP 120
Db 121 ASMLNPLOTSPSGGYPADATRPPLSLAQQALAPAMTPHQASLRYNTNGIDOKNTAM 180
   |||||||
Qy 121 ASMLNPLOTSPSGGYPADATRPPLSLAQQALAPAMTPHQASLRYNTNGIDOKNTAM 180
Db 181 LSPPHHTSTPQPYTEAMHINGYMSPYOAGGSPSYSOHHOSPPRHNNHHPMPT 240
   |||||||
Qy 181 LSPPHHTSTPQPYTEAMHINGYMSPYOAGGSPSYSOHHOSPPRHNNHHPMPT 240
Db 241 HENPEOVASPSIEDAPETKPTLVEPOSPKSPQNMKEELLRLVYNSPSEVERLKNKSG 300
   |||||||
Qy 241 HENPEOVASPSIEDAPETKPTLVEPOSPKSPQNMKEELLRLVYNSPSEVERLKNKSG 300
Db 301 ACSATNGPSRSKREKAAYIOETAGDEDEDESDSGTMSOGTTIYRRPKTERTAH 360
   |||||||
Qy 301 ACSATNGPSRSKREKAAYIOETAGDEDEDESDSGTMSOGTTIYRRPKTERTAH 360
Db 361 NLEKKYCSINDRIOQLVLLCGDAKLSKATLBRATIEHEVEHENQVLKHHEQMR 420
   |||||||
Qy 361 NLEKKYCSINDRIOQLVLLCGDAKLSKATLBRATIEHEVEHENQVLKHHEQMR 420
Db 421 KTLQNNRLPEPPIOTYTEXSARSPVSSPPRNERKSRMSTTPMKNGTDSGSKYTL 480
   |||||||
Qy 421 KTLQNNRLPEPPIOTYTEXSARSPVSSPPRNERKSRMSTTPMKNGTDSGSKYTL 480
Db 481 FAMLAVLLFNPGLAGSAIFSKAAAEAPLASPEHGVYIDDPDGTSTRTLFWESGIIN 540
   |||||||
Qy 481 FAMLAVLLFNPGLAGSAIFSKAAAEAPLASPEHGVYIDDPDGTSTRTLFWESGIIN 540
Db 541 MUYVWENLMTIYVVKLLIHGDPVQDMYSWOTFVTTREKARLNSGNLKDQRF 600
   |||||||
Qy 541 MUYVWENLMTIYVVKLLIHGDPVQDMYSWOTFVTTREKARLNSGNLKDQRF 600
Db 601 CECGLATLDRSLPSPGVDVSVSGWECVRHLNMLWIGRYIARRRSTKPVSVYCSHAO 660
   |||||||
Qy 601 CECGLATLDRSLPSPGVDVSVSGWECVRHLNMLWIGRYIARRRSTKPVSVYCSHAO 660
Db 661 TVLVYHEIHLQHLMTGTGTFEDTPEPSALTGLFMSLCAYNLAEAGASNDGLPRAVMAOI 720
   |||||||
Qy 661 TVLVYHEIHLQHLMTGTGTFEDTPEPSALTGLFMSLCAYNLAEAGASNDGLPRAVMAOI 720
Db 721 YTSASIOCLALPNLLAPFSGYFLRRARRHVRAPESHVSHLMTFHPATRKFSMDAR 780
   |||||||
Qy 721 YTSASIOCLALPNLLAPFSGYFLRRARRHVRAPESHVSHLMTFHPATRKFSMDAR 780
Db 781 LEHVLSKQKOLRFSGFEDEQLSPARIRTLKYLLSKLYOELVGDEIFTKNVERIL 840
   |||||||
Qy 781 LEHVLSKQKOLRFSGFEDEQLSPARIRTLKYLLSKLYOELVGDEIFTKNVERIL 840
Db 841 NONDRLEDVDVVDVSRLLVTTSTOCALITLNEKDSAFGTMISNRNGACCTWTHVLT 900
   |||||||
Qy 841 NONDRLEDVDVVDVSRLLVTTSTOCALITLNEKDSAFGTMISNRNGACCTWTHVLT 900
Db 901 CGIYMSRKNELAROHYSILRNCPPKILTDNLGLAVGHALCARICIDDRDSPKYSQYVC 960
   |||||||
Qy 901 CGIYMSRKNELAROHYSILRNCPPKILTDNLGLAVGHALCARICIDDRDSPKYSQYVC 960
Db 961 IHTKLSLESLELFTSTSSRASGVVSGIOEGT RMAWYIMNSLLDAMRSLFASKPYWTOS 1020
   |||||||
Qy 961 IHTKLSLESLELFTSTSSRASGVVSGIOEGT RMAWYIMNSLLDAMRSLFASKPYWTOS 1020
Db 1021 FFGQSTFSTLYOEAHYHAYIINGTRDGMRLFYELTCLMLGANQOATWSGRRVRSRK 1080
   |||||||
Qy 1021 FFGQSTFSTLYOEAHYHAYIINGTRDGMRLFYELTCLMLGANQOATWSGRRVRSRK 1080
Db 1081 MDVAVRGVSMRBSAOPDAFHLHTLVKLTSMDL 1113
   |||||||
Qy 1081 MDVAVRGVSMRBSAOPDAFHLHTLVKLTSMDL 1113

```

```

RESULT 2
ID Q24146 PRELIMINARY; PRT: 1113 AA.
AC Q24146:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
DE HLH106.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON S.
RX MEDLINE: 96165543.
RA THEOPOLD U., EKENGREN S., HULTMARK D.;
RT "HLH106, a Drosophila transcription factor with similarity to the
RT vertebrate steroid responsive element binding protein."
RL Proc. Natl. Acad. Sci. U.S.A. 93:1195-1199(1996).
DR EMBL: U38238; AAA97864.1; -.
DR HSSP: P36956; 1AM9.
DR FLYBASE: FBgn0015234; HLH106.
DR PFAM: PF00010; HLH; 1.
SQ SEQUENCE 1113 AA; 124513 MW; ED092420 CRC32;

Query Match 4.0%; Score 322; DB 5; Length 1113;
Best Local Similarity 24.6%; Pred. NO. 1,188-36;
Matches 192; Conservative 204; Mismatches 305; Indels 81; Gaps 68;

Db 261 SPANEOGVKVPINVOQKVEKRSANHAIERRISINDKINELKLVVGEQAKLMS 320
   |||
Qy 337 SEETMS-OGTIIYV-RPKT-E-RTAHNLEKKYCSINDRIOQLKVLGDEAKLSKS 392
Db 332 AVLRKSIDKINDIQONNDLKAELQRLQRELMARNGSVKLLQGTGPRASKRRRES 380
   |||
Qy 333 AVLRKSIDKINDIQONNDLKAELQRLQRELMARNGSVKLLQGTGPRASKRRRES 380
Db 381 QFTTDACTLPSPDESPPSLPMHSDISLPPSPYGSSTACSSGSSSNEEPLVPPSSM 440
   |||
Qy 441 ASRPVSSPPRR-NEKRS--RM-STTT-PMKN-GTRDGS-SKVTLPFAMLAIVLFPPI 493
Db 441 RGMATHSRGLCMFALVAVNPEKTEFLQRGHYSDNDLGMGSGORILSTDVSGEFAV 500
   |||
Qy 494 GLLA-GS-AI--FSKAA-AEAPIASPEHGVYIDDPD-GT-ST--RTL-FW-EG-ST-I 539
Db 501 WQSSWILNLTFTLMLGCVLTVYGDV-Q-LDAQT-LAYCOHROAFVFSOGSSQAY 557
   |||
Qy 540 -NMS-YVWVFNLMTIYVVKLLIHGDPVQDMYSWOTFVTTREKARLNSGNLKDQ 597
Db 558 AGYLCLMFGLSLPASRLCEYLQTTWQFLRFLRLWLGRLVSRSGGLFSMAARKQA 617
   |||
Qy 598 RKFCGLATLDRSLPSPGVDVSVSGWECVRHLNMLWIGRYIARRRSTKPVSVCS 657
Db 618 LASARELALNRLNQLDOLTN--GS--RGDMNGIMALLFASNAEVA--HNLLTPRETI 671
   |||
Qy 658 HAQTVLVYHEIHLQHLMTGTGTFEDTPEPSALTGLFMSLCAYNLAEAGASNDGLPRAV 717
Db 672 C-IHMTALBMKRSAPKMLQOEFARYMSRAROEGRFRATEOT-QELRMAFTAYGARY- 728
   |||
Qy 718 AOIYTSASIOCLALPNLLAPFSGYFLRRARRHVRAPESHVSHLMTFHPATRKFM 775
Db 729 CATVFTYDLSGSDGDFEFTRLRN-PCDPAHVITKQYREHLFFKSIQCLVGAGHSGGL 787
   |||
Qy 776 SDARKLEHVLSKQKOLRFSGFEDEQLSPARIRTLKYLLSKLYOELVG-GEIFTK 834
Db 788 PTSSVSGEAEOLQOQ-Q--HSGTIVSNV-LKTYSL-LKD-LMAD-ED--ERDTNV--VW 836
   |||
Qy 835 IVERILINDRLEDVDVVDVSRLLVTTSTOCALITLNEKDSAFGTMISNRNGACCTW 894
Db 837 MADVLETVAVHMLGSDTLAEQYGRIKQMPLOQOCCGNDHLPRALHVLAKMILKNN 896
   |||
Qy 895 WTHVLTGCIYMSRKNELAROHYSILRNCPPKIL--TUNLGLAVG-HA-LCARKICIDR 950

```

```

Db 897 GNALDKSKOLVNLCDSSVELQCLTVNRITDA-KGIKLFDLLTCDMLLETPTALME 955
OY 951 -DS-PK-VSOYVCITTKTSLESLRFLFSTSSRAGSVSGIOEGCTRMAYEWIMNSLDMR 1007
Db 956 LEHMMEDGFE-QY-PGE-VLEK-FQTDNLNLNIVENIPNAGSRITLYEAVCHLMGA 1011
OY 1008 -SNL-FASKPYWTQSFQSGSTFSTLYOEAYNHY-AIINGTRGDCWRFLFYELTCHMLNGA 1064
Db 1012 SP 1013
OY 1065 NP 1066

RESULT 3 PRELIMINARY: PRT: 403 AA.
ID 09WTN3:
AC 09WTN3:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE STEROL REGULATORY ELEMENT-BINDING PROTEIN-1 (SREBP-1) (FRAGMENT).
GN SREBP-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99161303.
RA INOUE J., SATO R.;
RT "A novel splicing isoform of mouse sterol regulatory element-binding
RT protein-1 (SREBP-1).";
RL Biosci. Biotechnol. Biochem. 63:243-245(1999).
DR EMBL: AB017337; BAA74795.1; -.
FT NON-TER 1 403
SQ SEQUENCE 403 AA: 41018 MW: 8577A8F CRC32:

Query Match 3.3%; Score 265; DB 11; Length 403;
Best Local Similarity 50.0%; Pred. No. 7,73e-26;
Matches 39; Conservative 18; Mismatches 21; Indels 0; Gaps 0;

Db 272 OSRGKRTAHNAIEKRRYSSINDKIVELKDLVGTGKAVLRKADYIRFLOHSNO 331
OY 351 RPKTERRTAHNLEKKYRCSINDRIQOLKVLGDEAKLSATLRAIHEIVEHEHO 410
Db 332 KIKQENILTRSAHKSKSL 349
OY 411 VKHHVEQMRKTLQNNRL 428

RESULT 4 PRELIMINARY: PRT: 144 AA.
ID 097676:
AC 097676:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE ADIPOCTE DETERMINATION AND DIFFERENTIATION-DEPENDENT FACTOR 1
DE (FRAGMENT).
GN ADD1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RA DING S.T., MERSMANN H.J.;
RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF102873; AAC78685.1; -.
DR HSSP: P36956; IAM9.
FT NON-TER 1 144
SQ SEQUENCE 144 AA: 15222 MW: 54B6A56 CRC32:

Query Match 2.9%; Score 229; DB 6; Length 144;

```

```

Best Local Similarity 55.9%; Pred. No. 2.80e-19;
Matches 33; Conservative 13; Mismatches 13; Indels 0; Gaps 0;

Db 85 GKALSSGSHQEKRTAHNAIEKRRSSINDKIVELKDLVGTGKAVLRKADYIR 143
OY 344 GTTIVRPKTERTAHNLEKKYRCSINDRIQOLKVLGDEAKLSATLRAIHEI 402

RESULT 5 PRELIMINARY: PRT: 498 AA.
ID P91527:
AC P91527:
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE SIMILARITY TO A HELIX-LOOP-HELIX MOTIF.
GN W02C12.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilitia; Rhabdilitida;
OC Rhabdilitia; Rhabdilitida; Rhabdilitidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAINES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,
RA GRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULLON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER R., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SANDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THERIERY-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMANN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RX MURRAY J., WOHLDMANN P.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U80815; AAB37997.1; -.
DR HSSP: P36956; IAM9.
DR PFAM: PF00010; HLH; 1.
SQ SEQUENCE 498 AA: 54391 MW: EAE8A806 CRC32:

Query Match 2.0%; Score 161; DB 5; Length 498;
Best Local Similarity 29.0%; Pred. No. 8.56e-08;
Matches 29; Conservative 34; Mismatches 30; Indels 7; Gaps 6;

Db 238 KKDINMIERRRYINRIKELGQMLPKNTSEDKLKGITLKASQYIRYLQKDEQA 297
OY 356 RRTAHNLEKKYRCSINDRIQOL-KVLL--CGDEAKLSATLRAIHEIVEHE-NOV 411
Db 298 KKTQ-QQ-QKSLESTPAHKYADRVKLEEMLARQGVQVPPS 335
OY 412 LKHVHEQMRKTLQNNRLPYEPI-QYTEYSARSPYESSPS 450

RESULT 6 PRELIMINARY: PRT: 110 AA.
ID 088368:
AC 088368:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE MICROPTHALMIA ASSOCIATED TRANSCRIPTION FACTOR (FRAGMENT).
DE MICROPTHALMIA.
OS Rattus norvegicus (Rat).

```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VARIOUS STRAINS;
RX MEDLINE: 98149870.
RA WEILBACHER K.N., HERSHEY C.L., TAKEMOTO C.M., HORSTMANN M.A.,
RA HEMESATH T.J., TASHJIAN A.H., FISHER D.E.;
RT "Age-resolving osteopetrosis: a rat model implicating microphthalmia
RT and the related transcription factor TFE3."
RL J. Exp. Med. 187:775-785(1998).
DR EMBL: AF029886; AAC26170.1; -.
DR HSSP: P22415; IAN4.
DR PFAM: PF00010; HLH; 1.
FT NON_TER 1 110
FT SEQUENCE 110 AA: 13155 MW; 7341AD76 CRC32;

Query Match 1.9%; Score 149; DB 11; Length 110;
Best Local Similarity 31.2%; Pred. No. 6.03e-06;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

DB 20 OKKDNNLIERRRRNINDRIKELGTLPKSNDDPMRNKGTILKASVDYIRKLOREQOR 79
OY 355 ERTAHNLEKKYRCISINDRIQOLKVLCC-GDEAKL--SKSATLRRAIEHIEVEHENOV 411
DB 80 AKD-LENROKLEHANR 95
OY 412 LKHNVQMKRTION-NR 427

RESULT 7
ID 070241 PRELIMINARY; PRT: 377 AA.
AC 070241;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE MITF (FRAGMENT).
GN WH.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98167856.
RA HODGKINSON C.A., NAKAYAMA A., LI H., SWENSON L.-B., OPDECAMP K.,
RA ASHER J.A. JR., ARNHEITER H., GLASER T.;
RT "Mutation at the anophthalmic white locus in Syrian hamsters:
RT haploinsufficiency in the Mitf gene mimics human Waardenburg syndrome
RT type 2."
RL Hum. Mol. Genet. 7:703-708(1998).
DR EMBL: AF020900; AAC15952.1; -.
DR HSSP: P22415; IAN4.
DR PFAM: PF00010; HLH; 1.
FT NON_TER 1 1
FT SEQUENCE 377 AA: 41728 MW; 83722925 CRC32;

Query Match 1.9%; Score 149; DB 11; Length 377;
Best Local Similarity 31.2%; Pred. No. 6.03e-06;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;
```

```
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE MITF PROTEIN (MICROPHthalmia-ASSOCIATED TRANSCRIPTION FACTOR).
GN MITF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKIN;
RX MEDLINE: 94348499.
RA TACHIBANA M., PEREZ-JORDAO L.A., NAKAYAMA A., HODGKINSON C.A., LI X.,
RA SCHNEIDER M., MIKI T., FEX J., FRANCKE U., ARNHEITER H.;
RT "Cloning of MITF, the human homolog of the mouse microphthalmia gene
RT and assignment to chromosome 3p14.1-p12.3."
RL Hum. Mol. Genet. 3:553-557(1994).
RN [2]
RP SEQUENCE OF 1-11 FROM N.A.
RX MEDLINE: 98160190.
RA WATANABE A., TAKEDA K., PLOPLIS B., TACHIBANA M.;
RT "Epistatic relationship between Waardenburg syndrome genes MITF and
RT PAX3."
RL Nat. Genet. 18:283-286(1998).
DR EMBL: Z29678; CAAB2775.1; -.
DR HSSP: P22415; IAN4.
DR PFAM: PF00010; HLH; 1.
KM DNA-binding.
SQ SEQUENCE 419 AA: 46938 MW; 5CE200FD CRC32;

Query Match 1.9%; Score 149; DB 4; Length 419;
Best Local Similarity 31.2%; Pred. No. 6.03e-06;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

DB 204 OKKDNNLIERRRRNINDRIKELGTLPKSNDDPMRNKGTILKASVDYIRKLOREQOR 263
OY 355 ERTAHNLEKKYRCISINDRIQOLKVLCC-GDEAKL--SKSATLRRAIEHIEVEHENOV 411
DB 264 AKE-LENROKLEHANR 279
OY 412 LKHNVQMKRTION-NR 427

RESULT 9
ID 073871 PRELIMINARY; PRT: 468 AA.
AC 073871;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE MITF.
GN CM19.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LECHORN; TISSUE-RETINAL PIGMENTED EPITHELIUM;
RX MEDLINE: 98133990.
RA MOCHII M., MAZAKI Y., MIZUNO N., HAYASHI H., EGUCHI G.;
RT "Role of Mitf in differentiation and transdifferentiation of chicken
RT pigmented epithelial cell."
RL Dev. Biol. 193:47-62(1998).
DR EMBL: D88363; BAA25648.1; -.
DR HSSP: P22415; IAN4.
DR PFAM: PF00010; HLH; 1.
FT NON_TER 1 1
FT SEQUENCE 468 AA: 52416 MW; 0686639D CRC32;

Query Match 1.9%; Score 149; DB 13; Length 468;
Best Local Similarity 31.2%; Pred. No. 6.03e-06;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

DB 253 OKKDNNLIERRRRNINDRIKELGTLPKSNDDPMRNKGTILKASVDYIRKLOREQOR 312
```



```

OY 335 ERRTAHNIIEKKYCSINDRIQOLKVLIC-GDEAKL--SKSATLRALIEHIEVEHENOV 411
Db 313 TKE-LENROKLEHANR 328
OY 412 LKHVHEQMRKTLQN-NR 427

RESULT 10
ID 008874 PRELIMINARY: PRT: 474 AA.
AC 008874: 008843: 008885: 060781: 060782;
DT 01-JUN-1998 (TReMBLrel. 05, Created)
DT 01-JUN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE MICROPHthalmia-ASSOCIATED TRANSCRIPTION FACTOR
DE (PUTATIVE TRANSCRIPTION FACTOR MI) (MICROPHthalmia-RELATED PROTEIN).
DE MITF OR MI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=LIVER:
RX MEDLINE: 93345026.
RA HODGKINSON C.A., MOORE K.J., NAKAYAMA A., STEINGRIMSSON E.,
RA COPELAND N.G., JENKINS N.A., ARNHEITER H.;
RT "Mutations at the mouse microphthalmia locus are associated with
RT defects in a gene encoding a novel basic-helix-loop-helix-zipper
RT protein."
RT Cell 74:395-404(1993).
RL [2]
RN SEQUENCE OF 238-285 FROM N.A.
RP STRAIN=C57BL/6; TISSUE=HEART;
RX MEDLINE: 94012591.
RA HUGHES M.J., LINGREL J.B., KRACKOWSKY J.M., ANDERSON K.P.;
RT "A helix-loop-helix transcription factor-like gene is located at the
RT mi locus."
RT J. Biol. Chem. 268:20687-20690(1993).
RN [3]
RP SEQUENCE OF 1-71 FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=C57BL/6; TISSUE=HEART;
RX MEDLINE: 9519171.
RA STEINGRIMSSON E., MOORE K.J., LAMOREUX M.L., FERRE-D'AMARE A.R.,
RA BURLEY S.K., SANDERS ZIMRING D.C., SKOW L.C., HODGKINSON C.A.,
RA ARNHEITER H., COPELAND N.G., JENKINS N.A.;
RT "Molecular basis of mouse microphthalmia (mi) mutations helps explain
RT their developmental and phenotypic consequences."
RT Nat. Genet. 8:256-263(1994).
RL [4]
CC -1- FUNCTION: ACTS AS A TRANSCRIPTION FACTOR.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH TFE3 OR TFE8 (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN, ISOFORMS 1 AND 2,
CC ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE. THE
CC SEQUENCE SHOWN HERE IS THAT OF ISOFORM 1.
CC -1- TISSUE SPECIFICITY: IN THE ADULT, EXPRESSED AT HIGH LEVELS IN THE
CC HEART, SKIN AND SKELETAL MUSCLE. BARELY DETECTABLE IN OTHER
CC TISSUES. IN THE EMBRYO, EXPRESSED IN DEVELOPING EYE, EAR, SKIN AND
CC HEART.
CC -1- DISEASE: DEFECTS IN THE MITF GENE ARE THE CAUSE OF MICROPHthalmia,
CC A CONDITION WHICH IS CHARACTERIZED BY ONE OR MORE OF THE
CC FOLLOWING: LOSS OF PIGMENTATION, REDUCED EYE SIZE, FAILURE OF
CC SECONDARY BONE RESORPTION, REDUCED NUMBERS OF MAST CELLS AND EARLY
CC ONSET OF DEAFNESS, AND WHICH GIVES RISE TO A NUMBER OF DIFFERENT
CC PHENOTYPES.
CC -1- SIMILARITY: TO OTHER MEMBERS OF THE MYC FAMILY OF HELIX-LOOP-HELIX
CC TRANSCRIPTION FACTORS.
DR EMBL: L22958; AAB47773.1; -
DR EMBL: Z23066; CA80600.1; -
DR EMBL: U19874; AAC52155.1; -
DR EMBL: U19875; AAC52156.1; -
DR HSSP: P22415; IAN4.
DR MGD: MGI:104554; Mtlf.
DR PFAM: PF00010; HLH; 1.

```

```

KW Nuclear protein; Transcription regulation; Alternative splicing;
FT DNA-binding; Phosphorylation; Disease mutation.
FT DOMAIN 201 217 ARG/LYS-RICH (BASIC).
FT DOMAIN 267 288 LEUCINE-ZIPPER (POTENTIAL).
FT DNA BIND 218 260 HELIX-LOOP-HELIX MOTIF.
FT VARSPIC 1 66 MTSRLIRLOOLMRQMOEORREDOCKLOAOEFOORVANS
FT OTPAINVSPTPLSPATOVEPEVLK -> MLEMETSHYO
FT (IN ISOFORM 2).
FT VARIANT 67 202 MISSING (IN MICROPHthalmia WHITE SPOT).
FT VARIANT 243 246 MISSING (IN MICROPHthalmia SPOTTED).
FT VARIANT 242 267 ACIPRESEBALAKEOKDNHLL -> V (IN
FT MICROPHthalmia EYELSS-WHITE).
FT VARIANT 267 267 I -> N (IN MICROPHthalmia WHITE).
FT VARIANT 271 271 R -> K (IN MICROPHthalmia OAK RIDGE).
FT VARIANT 272 272 MISSING.
FT VARIANT 277 277 D -> N (IN MICROPHthalmia VITILIGO).
FT VARIANT 318 474 MISSING (IN MICROPHthalmia CLOUDY-EYED).
SQ SEQUENCE 474 AA; 52991 MW; 8E464430 CRC32;

Query Match 1.9%; Score 149; DB 11; Length 474;
Best Local Similarity 31.2%; Pred. No. 6.03e-06;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

Db 259 OKDNHNIERRRPNDRIKELGTLPKSNDDPMRNKGTIKASYDYTRKLOREQOR 318
OY 335 ERRTAHNIIEKKYCSINDRIQOLKVLIC-GDEAKL--SKSATLRALIEHIEVEHENOV 411
Db 319 AKD-LENROKLEHANR 334
OY 412 LKHVHEQMRKTLQN-NR 427

RESULT 11
ID 075030 PRELIMINARY: PRT: 520 AA.
AC 075030.
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE A-TYPE MICROPHthalmia ASSOCIATED TRANSCRIPTION FACTOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE: 98321192.
RA AMAR S., FUSE N., YASUMOTO K., SATO S., YAJIMA I., YAMAMOTO H.,
RA UDONO T., DURLU Y.K., TAKAI M., TAKAHASHI K., SHIBAHARA S.;
RT "Identification of a novel isoform of microphthalmia-associated
RT transcription factor that is enriched in retinal pigment epithelium."
RL Biochem. Biophys. Res. Commun. 247:710-715(1998).
DR EMBL: AB006909; BAAS2288.1; -
DR HSSP: P22415; IAN4.
DR PFAM: PF00010; HLH; 1.
SQ SEQUENCE 520 AA; 58162 MW; F5D32118 CRC32;

Query Match 1.9%; Score 149; DB 4; Length 520;
Best Local Similarity 31.2%; Pred. No. 6.03e-06;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

Db 305 OKDNHNIERRRPNDRIKELGTLPKSNDDPMRNKGTIKASYDYTRKLOREQOR 364
OY 335 ERRTAHNIIEKKYCSINDRIQOLKVLIC-GDEAKL--SKSATLRALIEHIEVEHENOV 411
Db 365 AKE-LENROKLEHANR 380
OY 412 LKHVHEQMRKTLQN-NR 427

RESULT 12
ID 035889 PRELIMINARY: PRT: 1829 AA.
AC 035889.
DT 01-JAN-1998 (TReMBLrel. 05, Created)

```

Query Match	1.98;	Score 151;	DB 11;	Length 1829;
Best Local Similarity	23.4%;	Pred. No. 3.00e-06;		
Matches	59;	Mismatches 110;	Indels 15;	Gaps 14;

Db	1712	PP-ORNASTYKT	1722	:
		:: :		
QY	452	PRNERKRSRMST	463	:

```

RESULT      13
ID           021663
AC           021663; PRELIMINARY;
ID           021663; 002521; PRT; 281 AA.
DT 01-NOV-1996 (TIREMBLrel. 01, Created)
DT 01-NOV-1996 (TIREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TIREMBLrel. 08, Last annotation update)
DE SIMILAR TO HUMAN MAX INTERACTOR 1.
GN R03E9.1 OR MDL-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J.,
RA CRAXTON M., DEAR S., DU Z., DUBREIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KESHAM J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PAXSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SUTTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOHLIDAN P.,
RA "2.2 kb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."

```

Query Match	1.7%;	Score 139;	DB 5;	Length 281;
Best Local Similarity	26.9%;	Pred. No. 1.83e-04;		
Matches	32;	Conservative	36;	Mismatches 47;
			Indels	4;
			Gaps	4;

RESULT	14
ID	Q9MTW4
PRELIMINARY;	
PRT;	317 AA.

DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE TRANSCRIPTION FACTOR TFEC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE: 99156862.
RA REHLI M., DEN ELZEN N., CASSADY A.I., OSTROWSKI M.C., HUME D.A.;
RT "Cloning and characterization of the murine genes for bHLH-zip
RT transcription factors TFEC and TFEB reveal a common gene organization
RT for all Mrl subfamily members.";
RL Genomics 56:111-120(1999).
RL EMBL: AF077742; AAD24426.1; -.
SQ SEQUENCE 317 AA; 35143 MW; 54B4A438 CRC32.

Query Match	1.7%;	Score 135;	DB 11;	Length 317;
Best Local Similarity	27.2%;	Pred. No. 6.92e-04;		
Matches	25;	Conservative	29;	Mismatches 34;
			Indels	4;
			Gaps	3.

```
Db      110 QKKNHNLIERRRRYNINRYIKELGTLIPKSNDDPMRANKGTILKASVDYIKWLQKEQR 169
        :: |||||:: | :|| || :| : : : : :| : : :| :| :| :|
Ox      355 ERTAHNLTLEKKYKRSINDRILOOLKVLCC-GGEAKL--SKSATLRRAIHEIEVEHENOV 411
```

```

Db      170 ARE-LEHROKKLEHANKRLRIQLELEIQARA 200
      : ::::: || ||:
QY     412 LKHHEOMRKTLONNRLPYPEPIQYTEYSARS 443

```

RESULT	15	
ID	092UV8	PRELIMINARY;
AC	092UV8	PRT; 320 AA.
DT	01-MAY-1999	(TIREMBLrel, 10, Created)
DT	01-MAY-1999	(TIREMBLrel, 10, Last sequence update)
DT	01-MAY-1999	(TIREMBLrel, 10, Last annotation update)

480 251 / 160
09

THIS PAGE BLANK (USPTO)


```

CC CC IS A LIM DOMAIN-CONTAINING PROTEIN. SRD PHENOTYPES ARE RESISTANT
CC CC TO STEROL BIOSYNTHESIS REPRESSION BY STEROLS.
CC CC - SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC at the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sb.ch/announce/
CC CC or send an email to license@isb-sb.ch).
CC CC -----
DR DR EMBL: U12330; AAA74141.1; .
DR DR EMBL: U12329; AAA74140.1; ALT_TERM.
DR DR EMBL: U22819; AAA85719.1; ALT_TERM.
DR DR EMBL: U22818; AAA85718.1; ALT_TERM.
DR DR HSSP: P36956; 1AM9.
DR DR PFAM: PF00010; HLH; 1.
DR DR PROSITE: PS00038; HELIX_LOOP_HELIX; FALSE NEG.
KW KW Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;
KW KW Endoplasmic reticulum; Polymorphism; Chromosomal translocation.
FT FT DOMAIN 1 50
FT FT DOMAIN 51 120
FT FT DOMAIN 121 244
FT FT BINDING 95 421
FT FT TO STEROL REGULATORY ELEMENT-1 (BY
FT FT SIMILARITY).
FT FT DNA_BIND 328 341
FT FT DOMAIN 342 379
FT FT DOMAIN 378 399
FT FT SITE 460 461
FT FT LEUCINE-ZIPPER (POTENTIAL).
FT FT BEACPOINT FOR TRANSLLOCATION TO FORM
FT FT SREBP-2 FUSION PROTEINS IN SRD
FT FT PHENOTYPES.
FT FT CLEAVAGE (BY APOPAIN AND CASPASE-7) (BY
FT FT SIMILARITY).
FT FT TRANSMEM 480 500
FT FT TRANSMEM 532 552
FT FT DOMAIN 589 593
FT FT DOMAIN 857 860
FT FT VARIANT 493 493
FT FT SEQUENCE 1139 AA; 123655 MW; E81C2778BEP02653 CRC64;
SO SO
Query Match 5.5%; Score 443; DB 1; Length 1139;
Best Local Similarity 25.5%; Pred No.1,15e-63;
Matches 193; Conservative 208; Mismatches 280; Indels 75; Gaps 56;
Db 324 PREGERTTHNIIEKRYSSINDKIIELDWLGTDAKKHKSGVLRAKDAIDYIKYLQOVNH 383
|| |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 352 PK-TERRAHNNIEKKYRCSINDRIQQLKVLGCGDAKSKATLRALIEHI---EEVGH 407
Db 364 KLRQENWVLKLANQKNKLLKIGDLSVDVDLKIEDFNQNVLLMSPASDSGSOAGFS 443
|| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy 408 ---ENQVYLK-HHYEOMKRTLQNNRLPYPE-PIQYTERSARSPVSSPPRNERK--R 458
Db 444 PYSIDSEFGSPILLDDAKVDEDPDPPVLAAGMDRSKILLCYTLFGLSPSLTSLQNG 503
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy 459 S-RM-S-TTTPK-KNGT-RD-GSSKVTYLF--A-MLLAVLIENFDIGLAGSAIFSKAA 506
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 504 AHDQHPYSGSGRSVLSLESGSG-GWF-D-WMMPTLLMLVNGVVLVSVFKALVHGEP 560
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy 507 AEAPVAPSPPE-HGRVYIDDPDGTSTRLEFEGSITNKSYYWNIMILYVVKLIHGDP 565
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 561 VIRPHTSSVTFEWRRRKQADLDLARGDEFAAANLOTCLSVLGRALPTSRDLACSLSNW 620
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy 566 VQDFMSVSMQGTIVTTREKARAELENSGNLMDAQRKPOECPLATIDRLSPGCVDSVSGNE 625
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 621 VTRYSQIKRLVRLWIKKYFQWRKATPATYAGFEDEAKSSASDAALAYHRLQHL---IT 677
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy 626 CVRRH-LLN-WL-WIGRYTARRRRST-TKPVSVV-CRSHAOYAVL-YEHIQLHMGIT 677
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 678 GKLPV-A-GSTC-SDVHMAKAVNLAECA---EKKIPSTLIEHLITLAAGLTGTGGGKRG 731
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy 678 GNFEDTPEPSALTGILPMSICAVNLAECAGASMDGLPRAVMAQITISASIQCRALPNTLTA 737

```

ID	RESULT	2	STANDARD:	PRT: 1141 AA.
AC	012772:			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	STEROL REGULATORY ELEMENT BINDING PROTEIN-2 (SREBP-2) (STEROL REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 2).			
GN	SREBP2 OR SREBP2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 94089681.			
RA	Hua X., Yokoyama C., Wu J., Briggs M.R., Brown M.S., Goldstein J.L., Wang X.;			
RT	"SREBP-2, a second basic-helix-loop-helix-leucine zipper protein that stimulates transcription by binding to a sterol regulatory element.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 90:11603-11607(1993).			
RT	[2]			
RP	SEQUENCE OF 91-109.			
RX	MEDLINE: 94006541.			
RA	Yokoyama C., Wang X., Briggs M.R., Admon A., Wu J., Hua X., Goldstein J.L., Brown M.S.;			
RT	"SREBP-1, a basic-helix-loop-helix-leucine zipper protein that controls transcription of the low density lipoprotein receptor gene.";			
RL	Cell 75:187-197(1993).			
RN	[3]			
RP	CLEAVAGE AT ASP-468 BY CASPASES.			
RX	MEDLINE: 96224303.			
RA	Pai J.-T., Brown M.S., Goldstein J.L.;			
RT	"Purification and cDNA cloning of a second apoptosis-related cysteine protease that cleaves and activates sterol regulatory element binding proteins.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 93:5437-5442(1996).			
CC	-1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE STEROL REGULATORY ELEMENT 1 (SRE-1) (5'-ATCACCCGAC-3') FOUND IN THE FLANKING REGION OF THE LDLR AND HMG-COA SYNTHASE GENES.			
CC	-1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER BHL PROTEIN.			
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANES OF THE NUCLEAR ENVELOPE AND ENDOPLASMIC RETICULUM. RELEASED INTO THE NUCLEUS UPON PROTEOLYTIC CLEAVAGE (BY SIMILARITY).			
CC	-1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN ADULT AND FETAL TISSUES.			

```

CC -1- PTM: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY
CC CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS
CC AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE
CC PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,
CC INDEPENDENT OF STEROL LEVELS.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U02031; AAA50746.1; -.
DR HSSP: P36956; 1AM9.
DR PFM: 600481; -.
DR PFM: PF00010; HLH: 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; FALSE NEG.
KW Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;
KW Endoplasmic reticulum.
FT DOMAIN 1 50 TRANSSCRIPTIONAL ACTIVATION (ACIDIC).
FT DOMAIN 51 124 GLY/PRO/SER-RICH.
FT DOMAIN 125 246 GLY/PRO/SER-RICH.
FT DOMAIN 95 423 BINDS TO STEROL REGULATORY ELEMENT-1 (BY
FT SIMILARITY).
FT DNA_BIND 330 343 BASIC DOMAIN.
FT DOMAIN 344 381 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 380 401 LEUCINE-ZIPPER (POTENTIAL).
FT SITE 468 469 CLEAVAGE (BY APOPAIN AND CASPASE-7).
FT TRANSMEM 482 502 POTENTIAL.
FT TRANSMEM 534 554 POTENTIAL.
FT DOMAIN 591 595 POLY-ALA.
SQ SEQUENCE 1141 AA: 123673 MW: DE42B9C16C832B CRC64:
Query Match 5.3%; Score 421; DB 1; Length 1141;
Best Local Similarity 25.1%; Pred. No. 5,596-59;
Matches 190; Conservative 210; Mismatches 280; Indels 77; Gaps 59;
Db 326 KPEEERTTHIIKRRRSSINDKIELKLDVMTGDKMHRSGVLRKAIDYIKYIQGVNH 385
QY 352 PK-TERRAHNLIEKKYKCSINDRIQQLKVLICGDEAKLSKSLRAIEHI---EVEEH 407
Db 386 KLRQENWLKLANQKNLKGIDLSLDNEVDKIEDFNQNVLLMSPASSSSQAGFS 445
QY 408 ----ENOVLK--HHVEQRRKTLQNNRLPYE-PIQYTESARSPVSSPSPRNERK--R 458
Db 446 PYSIDSEPGSPLLDDAKKDEPDSPVALGMVDRSRILLVLTFLCLSFNPLTSLQWG 505
QY 459 S-RM-S-TTITM-KNGT-RD--GSSKVTL-F--A-MLAVLIENPILLGASAFISKAA 506
Db 506 AHSDDQPHSSGSGSVLSFESGSG-GWF-D-WMPPTLLMLVNGVIVSVFKLLHGE 562
QY 507 AEADIASFE-HGRVIDPDGTSTRTLLFWESIIIMSYWFNIMITYVWKLIIHSDP 565
Db 563 VIRPHSRSVTFEMRRKQADLDIARGDFAAANQTCIAYLGRALPTSRIDLACLSLWN 622
QY 566 VQDFNSVSWQTFVITREARAEIENSGNLKDAORKCECLATIDRSLPSPGVDSVSGWE 625
Db 623 VIRSYLQKRLVRLMLLKVFQCRATPATTEAGFEDEAKTSARDAALAYRLHQLH---IT 679
QY 626 CVRH-LIN-WL--WI-GR-YIARRRRSTTKPV-SYVGRSHQOTAVL-YHEIHQHLMIIT 677
Db 680 GLP-A-GSAC-SDVHMLCAVNAIECA---EKKIPSTIVEIHLTAAMGLKTRCGKLG 733
QY 678 GNFDYTPYSALTLGFLMSICAVNIAEAAGASNDGJPRAVMQIYISAIQCRALPNDLA 737
Db 734 -FLASYFLSPKOS-LC-GPEHSAVPDSLRLCHPLGOKFEMRSMYSVS-AAKES-L-Y- 786
QY 738 PFFSGYFLRRARRHVRAPSHS-VSHLL-WIHPATRKFMDSAKLEHVLSSKOKOLRF 795

```

```

Db 787 C-AQRNPADPIAQVHOAFCKNLERALESLVYKPOAK-KKAGDQ-EESCEFFSALEYL-- 841
QY 796 SFVEDQSLPARILRTLLKVLKSLVQELGCGDEIFKNVERILINDRDDEVDVDDV 855
Db 842 -KLHSP-VDSVGVMSPLSSVLSK--ALGPILICRWMTSATVAIWSLQGDAAVRS 897
QY 856 SRLVITISTQCAAILITNKEDSAKFGTWSIRNGACCTWMTHTVLTGCIYMSNKNELARQ 915
Db 898 HETKWERIKALEVETESPIYKAIFHA-CRAMHASTLPKADQOSSF-C-HCERA-SG-HL 952
QY 916 HYSLRNCPKRI-LTDN-LGLAVGHALC-AKICIDDRDSKVSYQYCIHTRKKSLESLRL 972
Db 953 WSSLNVSQGISDPALNHYVOLLTDLL-LSTALMOKQASASQAV-GETYHASCAGELAG 1010
QY 973 FSTSRAGGVV-SGIQECTRMAYEWIMNSLDA-WRSNLFASPYMWTQSGQSTFSTL 1030
Db 1011 FQDRLGSLRIKASHFRPAYRKVFLHEATVRLMAGGSP 1047
QY 1031 YQEAVNHYA-IINGTRGDCWRLFVYELTCRMLNGANP 1066
RESULT 3
ID SREL_CRIGR STANDARD: PRT: 1133 AA.
AC 060416:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, last sequence update)
DT 15-FEB-2000 (Rel. 39, last annotation update)
DE STEROL REGULATORY ELEMENT BINDING PROTEIN-1 (SREBP-1) (STEROL
DE REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 1).
GN SREBP1 OR SREBP1.
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94274723.
RA Sato R., Yang J., Wang X., Evans M.J., Ho Y.K., Goldstein J.L.,
RA Brown M.S.;
RT "Assignment of the membrane attachment, DNA binding, and
RT transcriptional activation domains of sterol regulatory
RT element-binding protein-1 (SREBP-1).";
RL J. Biol. Chem. 269:17267-17273(1994).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE STEROL
CC REGULATORY ELEMENT 1 (SRE-1) (5'-ATCACCACGAC-3') FOUND IN THE
CC FLANKING REGION OF THE LDLR GENE AS WELL AS OTHER GENES.
CC ADD1/SREBP1 HETERODIMER HAS DUAL SEQUENCE SPECIFICITY, BINDING TO
CC BOTH AN E-BOX MOTIF (ATCAGCTCA) AND TO SRE-1.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE OF THE NUCLEAR
CC ENVELOPE AND ENDOPLASMIC RETICULUM. RELEASED INTO THE NUCLEUS UPON
CC PROTEOLYTIC CLEAVAGE (BY SIMILARITY).
CC -1- PTM: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY
CC CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS
CC AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE
CC PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,
CC INDEPENDENT OF STEROL LEVELS.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U09103; AAA20085.1; -.
DR HSSP: P36956; 1AM9.
DR PFM: PF00010; HLH: 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX: 1.

```

[illegible]

```

RC      TISSUE=ADIPOCYTE.
RX      MEDLINE: 93303269.
RA      Tortoreo P., Kim J.B., Graves R.A., Spiegelman B.M.:
RT      "Add1: a novel helix-loop-helix transcription factor associated with
RL      Mol. Cell. Biol. 13:4753-4759(1993).
RN      [2]
RP      DNA-BINDING.
RX      MEDLINE: 95257939.
RA      Kim J.B., Spotts G.D., Halvorsen Y.D., Shin H.M., Ellenberger T.,
RA      Towle H.C., Spiegelman B.M.:
RT      "Dual DNA binding specificity of ADD1/SREBP1 controlled by a single
RL      amino acid in the basic helix-loop-helix domain.".
RL      Mol. Cell. Biol. 15:2582-2588(1995).
CC      -1- FUNCTION: REGULATED DURING BOTH ADIPOCYTE DETERMINATION AND
CC      DIFFERENTIATION. ADD1/SREBP1 HEMEROIDIN HAS DUAL SEQUENCE
CC      SPECIFICITY. BINDING TO BOTH AN E-BOX MOTIF (ATCACGGA) AND TO
CC      SRE-1 (5'-ATACCCCGC-3').
CC      -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC      BHLH PROTEIN.
CC      -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS: SREBP-1A AND SREBP-1C/ADD1
CC      (SHOWN HERE): ARE PRODUCED BY ALTERNATIVE SPLICING.
CC      -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BROWN ADIPOSE
CC      TISSUE.
CC      -1- PTM: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY
CC      CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS
CC      AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE
CC      PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,
CC      INDEPENDENT OF STEROL LEVELS (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC      TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC      -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. MANY PROBABLE
CC      FRAMESHIFTS WERE CORRECTED FROM POSITION 878 ONWARD.
CC      -----
CC      THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC      BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC      THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC      USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC      MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC      ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sib.ch/announce/
CC      OR SEND AN EMAIL TO license@isb-sib.ch).
CC      -----
DR      EMBL: L16995. -: NOT ANNOTATED CDS.
DR      PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW      Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;
KW      Endoplasmic reticulum; Activator; Apoptosis; Alternative splicing.
FT      BINDING 67 385
FT      SIMILARITY.
FT      DOMAIN 125 128
FT      DNA_BIND 293 306
FT      DOMAIN 307 344
FT      DOMAIN 343 364
FT      SITE 427 428
FT      TRANSMEM 454 474
FT      TRANSMEM 513 533
FT      NON_TER 1023 1023
SO      SEQUENCE 1023 AA; 108571 MW; 2593DF46B2A11C8F CR664;

Query Match 4.7%; Score 375; DB 1; Length 1023;
Best Local Similarity 30.2%; Pred. No. 2,656-49;
Matches 155; Conservative 125; Mismatches 180; Indels 53; Gaps 37;

Db 282 GKALGSAOSRBEKRRANHAIEKRRYRSINDKIVELKDLVGTAEKLNSAYLRAIDYIR 341
QY 344 GTTIVRERKERRRANHLIKRYKRSINDRIIOQLKLCDEALKEASATLRAIRIE 403
Db 342 FLOHNSOKLKOENITLRAHKSRLKLDVSRGSGGGGVDVMECKKPVVETLTPPPSDA 401
QY 404 EYEHENQVLKHHVDMOKRTLDN--R-LPY-EPLOYEYS--RSP--VES-SRSP-R 453
Db 402 GSPSPSLFSLKSGSSGGSDESDPAFEDNOYKRAQLRPSHSGMLDPSRLALCVLF 461

```


454 NERKSR-MSTTP-MKNGTRGSSKVTTLFA-MLL-AV-LIENPGLLAGS--A--IFS- 503
 Db 462 LCLTCNPALSLFGWGLTPSDASGVHRSSGRSMLEASRDGSMWTOMLPLVWLANGLL 521
 504 KAAAEPAIASPFHGRVI-DDPDGT--ST-RILE----WEGSI-IN-KS-YWVNIIML 551
 Db 522 VLACLALFVYGVPVTRPHSGPAVHFWRHRRKQADLDLARGDFAQAQOIMLALQALGRPL 581
 552 IIVYVVKLLIHGDPPVODFMSVSWQTFVTTREKARAEINSQNLKDAQRKFCCECIATIDRSL 611
 Db 582 PPSNLDLACSLMLNIVRHILQRLWGRWMLAGQAGQORRYRLKKDARASRDAAVYVYHKL 641
 612 PSPGVDSFVSQWECVRLHMLWMLIGRIYLRARRSTTK--PVSVCGRSHAQ-TAVLYHET 668
 Db 642 HOLHMG--KYTGCH-LVA-SNL--ALSALNLAEACAG--DAISMATLAEIYVAALRY 691
 669 HOLHMGITNGNEDITYEPASALIGLEFMSLCVNLAEAGASNDGLPAAVQAQIITSISIQ 728
 Db 692 KTSLEPRAL-HLITREFLSSAR-QACLAQSGAVPLAMQWICHPYGRHFVYDGMVAHVH-AP 748
 729 RALPLMLAPFSGYFLRRARRHVRAPHSVS-HLIMIFPATRKREMSDARKLEHYLSS 787
 Db 749 -QESL-Y-S-VAGNPVDPPLAQVTRLFCHEHLER 777
 788 KOKQLRFGSFEDEQLSPILARITTLKXYLLSK 820
 RESULT 5 SRE1 HUMAN STANDARD: PRT: 1147 AA.
 AC P36956:
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE STEROL REGULATORY ELEMENT BINDING PROTEIN-1 (SREBP-1) (STEROL
 DE REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 1).
 GN SREBP1 OR SREBP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE: 94005541.
 RA Yokoyama C., Wang X., Briggs M.R., Admon A., Wu J., Hua X.,
 RA Goldstein J.L., Brown M.S.;
 RT "SREBP-1, a basic-helix-loop-helix-leucine zipper protein that
 RT controls transcription of the low density lipoprotein receptor
 RT gene".
 RL Cell 75:187-197(1993).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 319-394.
 RX MEDLINE: 98298442.
 RA Partridge A., Belisolell L., Ferre-D'Amaré A.R., Burley S.K.;
 RT "Co-crystal structure of sterol regulatory element binding protein 1a
 RT at 2.3-A resolution".
 RL Structure 6:661-672(1998).
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE STEROL
 CC REGULATORY ELEMENT 1 (SRE-1) (5'-ATCACCACAC-3'), FOUND IN THE
 CC FLANKING REGION OF THE LDLR GENE AS WELL AS OTHER GENES.
 CC ADD1/SREBP1 HETERODIMER HAS DUAL SEQUENCE SPECIFICITY, BINDING TO
 CC BOTH AN E-BOX MOTIF (ATCAGTGA) AND TO SRE-1.
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
 CC BHLH PROTEIN.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANES OF THE NUCLEAR
 CC ENVELOPE AND ENDOPLASMIC RETICULUM. RELEASED INTO THE NUCLEUS UPON
 CC PROTEOLYTIC CLEAVAGE.
 CC -1- ALTERNATIVE PRODUCTS: SEVERAL ISOFORMS DIFFERING IN THE N- OR C-
 CC TERMINI ARE PRODUCED BY ALTERNATIVE SPLICING. THE SEQUENCE SHOWN
 CC IS THAT OF SREBP-1A.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A WIDE VARIETY OF TISSUES, MOST
 CC ABUNDANT IN LIVER AND ADRENAL GLAND. IN FETAL TISSUES LUNG AND
 CC LIVER SHOWS HIGHEST EXPRESSION.
 CC -1- PIM: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY
 CC CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS

CC AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE
 CC PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,
 CC INDEPENDENT OF STEROL LEVELS.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>
 CC or send an email to license@isb-stb.ch).
 CC -----
 CC EMBL: U00968; CAB34682.1; -
 DR PDB: 1AM9; 01-JUL-98.
 DR TRANSFAC: T01556; -
 DR MIM: 184756; -
 DR PFAM: PF00010; HLH; 1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
 KW Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;
 KW Endoplasmic reticulum; Activator; Apoptosis; 3d-structure;
 FT DOMAIN 61 178
 FT BINDING 92 416
 FT TO STEROL REGULATORY ELEMENT-1 (BY
 FT SIMILARITY).
 FT BASIC DOMAIN.
 FT DNA_BIND 323 336
 FT DOMAIN 337 374
 FT DOMAIN 373 394
 FT DOMAIN 427 462
 FT SITE 460 461
 FT CLEAVAGE (BY APOPAIN AND CASPASE-7) (BY
 FT SIMILARITY).
 FT TRANSMEM 488 508
 FT TRANSMEM 548 568
 FT POTENTIAL.
 SQ SEQUENCE 1147 AA; 121644 MW; 58ED093A2371FE3E9 CRC64;
 Query Match 4.5%; Score 358; DB 1; Length 1147;
 Best Local Similarity 25.9%; Pred. No. 8,95e-46;
 Matches 195; Conservative 195; Mismatches 299; Indels 72; Gaps 54;
 Db 87 LSGPQAAPSPPLPPAPPLPKMYSPMAFSPGPKIESVP-LSTLQTPPOPLGALL 145
 115 LSG-EGPASHLNPLOPSP-PSGGYPADAYRPLSLAQILAAPAMPHQAAISFVNTGID 172
 Db 146 POSFP-AP-APQFQSTPYLQYRSPPGCFSTGSPGNTQQLPLGULPASPQVPPVSLHT 203
 173 QKNFTAMLSPHHTSMISQPYTEAMGHTNGYSPYDQAQGSQSPSYQHQSPPHHH 232
 Db 204 QVQSVVP-QQLLTVAAPRAAPVTTVTTSIQIOQVPLLOPHFIKADSL-LTAMK-TDGA 260
 233 HHHPMKTHIENEQVYASPIEDAPETKPTHLVPEOSPKSPQMKKEILLRLNMSPSVEV 292
 Db 261 TYKA-AGISPLVSGTIVTQGPLPLVSGGTLTATVPLVDAEKPLINRLAAGSKAPASA 318
 293 RLKNNKSCASATNGPSRFSREKAAKIVIDETAEGBED-EDDEDSQSGETMSGGTTIVR- 350
 Db 319 QSRGEKRTAHNAIEKRYBSINDKILKDLVYGTEAKNKSAVLRKADYTRFLOHSQ 378
 351 RKTERRTAHNIEKRYBSINDRIQQLKVLGCGEAKLSKATLRRALTEHIEVEEHENO 410
 Db 379 KIKQENLSRTVAHRSKSLKDLSVAGSGGNTDVLMEGKTEVEPTLPPPDAGSPFGS 438
 411 VLKHYEOMKRTLQNNR-L-PY-PE--PQYTE-Y-SKRSVESPSPRNE--R-KR 458
 Db 439 SPLSLGRSGSGSGSDSEPDSPVFEEDSKAPPEORPSLSRGMIDRSLALCTIVFLCL 498
 459 SRMS-TT-TPMNGT-ROSSKAVTLF--AMLAVALI--FNPIGLLA-G--S--A-IF--- 502
 Db 499 SCNPPLASLIGANGLPSPSPTTSVYHSPGRNVLTGSRDQPGQAQMLLP-PVWLLGLLV 557
 503 S-KA-AA--EAP-IASPFHGRVIDDP--D--GTSTRT-LFWEKGSITNMSYVWVNIIML 552
 Db 558 LVSIVLFFYGGPVRPHSGPAVHFWRHRRKQADLDLARGDFAQAQOIMLALQALGRPL 617

QY 553 IYVVVLLHIGDPVODFMVSMOTFTTRKARAEINSGNLKDAQCFCECLATLDRSLP 612
DB 618 TSHLDIACSLMNLIRHILORLWGMVIRAGAGLODDCARPDASABDALVYHKLH 677
QY 613 SPQVDSVFSVGMVCHRLMWMIGRIYARRRSTKPVSV-V-CRSHQTAFL-YHEIH 669
DB 678 QLTHTG-K-H--T-G-GHLLATNALALNLEACAG---DAVSATVLAETIYVAALRYK 727
QY 670 QHLMGITGNEFDTEPRLTGLFMSLCANVLEAGASDGLPRAVMQIYISASIQCR 729
DB 728 TSLPRAL-HELTREFLSSAR-QACLAQSGSVPRAMQMLCPVCHRRFVQGDW-S-VLSTP 783
QY 730 LALPNLAEFFSGYFLRRARRHVRAPHSVSHL-WIFHPARKFMSAKRLEHVLSSK 788
DB 784 WSLT-Y-SLAGN-PVDPLAQVTLFHEHLER 812
QY 789 OKOLRFGSFEVDEOLSPLRIRITTLKAVYLISK 820

RESULT 6
AC TFE3 MOUSE STANDARD: PRT: 446 AA.
ID 064092:
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 15-FEB-2000 (rel. 39, Last annotation update)
DE TRANSCRIPTION FACTOR E3 (FRAGMENT).
GN TFE3 OR TCFE3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92123207.
RA Roman C., Materna A.G., Cooper C., Ariandi S., Blain S., Ward D.C.,
Calame K.;
RT "TFE3, an X-linked transcriptional activator containing basic helix-
loop-helix and zipper domains, utilizes the zipper to stabilize both
RT DNA binding and multimerization."
RL Mol. Cell. Biol. 12:817-827(1992).
CC -1- FUNCTION: POSITIVE-ACTING TRANSCRIPTION FACTOR THAT BINDS TO THE
IMMUNOGLOBULIN ENHANCER MUE3 MOTIF. IT BINDS ALSO VERY WELL TO A
CC USP/MLF SITE. BINDING OF TFE3 TO DNA INDUCES DNA BENDING.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
BHLH PROTEIN.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: S76673; AAB21130.1; -
DR HSSP: P36956; IAW9.
DR MGD: MGI:98511; TCFE3.
DR PFAM: PF00010; HLH.1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX.1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT NON_TER 1
FT DOMAIN 133 144 STRONG TRANSCRIPTION ACTIVATION DOMAIN
FT 217 (POTENTIAL).
FT DNA_BIND 217 232 BASIC DOMAIN.
FT DOMAIN 233 273 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 282 303 LEUCINE-ZIPPER (POTENTIAL).
SO SEQUENCE 446 AA; 47891 MW; 79115373AD7F131E CRC64;

Query Match 1.88; Score 143; DB 1; Length 446;
Best Local Similarity 31.28; Pred. No. 1.82e-05;

Matches 24; Conservative 26; Mismatches 22; Indels 5; Gaps 4;
DB 219 OKKDNHLEERRRRNINDRIKELTLPKSNDEPRNRKGIILKASYDIKLOEQR 278
QY 355 ERTAHNLEIKRYKCSINDRIQOLKVLVC-GDEAKL--SKSATLRALRHEIEVEHENY 411
DB 279 SKD-LESQRSLQOANR 294
QY 412 LKHVYQMRKTL-QNNR 427

RESULT 7
AC BFL_HUMAN STANDARD: PRT: 477 AA.
ID P55315:
DT 01-OCT-1996 (rel. 34, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 01-OCT-1996 (rel. 34, Last annotation update)
DE TRANSCRIPTION FACTOR BF-1 (BRAIN FACTOR 1) (BFL) (HFRL).
GN FKHL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 95048332.
RA Murphy D.B., Wiese S., Burfeind P., Schmundt D., Mattei M.-G.,
Schulz-Schaeffer W., Thies U.;
RT "Human brain factor 1, a new member of the fork head gene family,"
RL Genomics 21:551-557(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 95322450.
RA Wiese S., Murphy D.B., Schlung A., Burfeind P., Schmundt D.,
Schulze V., Mattei M.-G., Thies U.;
RT "The genes for human brain factor 1 and 2, members of the fork head
RT gene family, are clustered on chromosome 14q."
RL Biochim. Biophys. Acta 1262:105-112(1995).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ESTABLISHMENT OF THE
REGIONAL SUBDIVISION OF THE DEVELOPING BRAIN AND IN THE
DEVELOPMENT OF THE TELENCEPHALON. SEQUENCE-SPECIFIC DNA-BINDING
CC PROTEIN WITH A DISTINCT BINDING SPECIFICITY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: X74142; CAA52239.1; -
DR MIM: 164874; -
DR PFAM: PF00250; Fork_head.1.
DR PRINTS: PR00053; FORKHEAD.
DR PROSITE: PS00657; FORK_HEAD_1.1.
DR PROSITE: PS00658; FORK_HEAD_2.1.
DR PROSITE: PS50039; FORK_HEAD_3.1.
KW Transcription regulation; DNA-binding; Nuclear protein;
Developmental protein.
FT DOMAIN 33 57 HIS-RICH.
FT DOMAIN 58 80 PRO-RICH.
FT DOMAIN 72 75 POLY-GLN.
FT DOMAIN 99 102 POLY-LEU.
FT DNA_BIND 168 259 FORK-HEAD.
SO SEQUENCE 477 AA; 51340 MW; 71CFD0BD069C9AD5 CRC64;

Query Match 1.88; Score 145; DB 1; Length 477;
Best Local Similarity 29.18; Pred. No. 9.01e-06;
Matches 37; Conservative 27; Mismatches 55; Indels 8; Gaps 8;

Db	39	HHNSHHPDHHHHHHHHHHHHHHPPRRPAPQPPRRPAAAOQOOPRRPRLPQAGGAQNSDKEGPQ	98
Oy	223	HHOSPPRRHHHHHHHPMK.IHNEPVOVASPIDADEPTKTL-VEPOSPKSPONKKELLR	281
Db	99	LL-L-PPRDHHRPSPGAAGCCRGCGELGPRVGPDEKEKGAAGGEEKKGAFGSGKDOEGC	157
Oy	282	LLVNSPSEVERL-KNKKSGACS-ATN-GP-SRS-KEKAATVIQET-AEGDEDEDEDS	335
Db	158	KEGEKN 164	
Oy	336	DSGETMS 342	
RESULT	8		
ID	TFE3_HUMAN	STANDARD:	PRT: 743 AA.
AC	P19532; Q99964; Q92757; Q92758;		
DT	01-FEB-1991 (Rel. 17, Created)		
DT	15-JUL-1999 (Rel. 38, Last sequence update)		
DT	15-FEB-2000 (Rel. 39, Last annotation update)		
DE	TRANSSCRIPTION FACTOR E3.		
GN	TFE3.		
OS	Homo sapiens (Human).		
OC	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
RN	SEQUENCE OF 1-219 FROM N.A., AND CHROMOSOMAL TRANSLOCATION.		
RP	MEDLINE: 97140324.		
RX	Weerman M.A.J., Wilbrink M., Geurts van Kessel A.;		
RA	"Fusion of the transcription factor TFE3 gene to a novel gene, PRC3,		
RT	in t(X;1)(p11;q21)-positive papillary renal cell carcinomas.",		
RL	Proc. Natl. Acad. Sci. U.S.A. 93:15294-15298(1996).		
RN	[2]		
RC	SEQUENCE OF 149-743 FROM N.A., AND CHROMOSOMAL TRANSLOCATION.		
RP	TISSUE=MONOCYTES.		
RX	MEDLINE: 97026295.		
RA	Sidhar S.K., Clark J., Gill S., Hamoudi R., Crew A.J.,		
RA	Gwilliam R., Ross M., Linehan W.M., Birdsall S., Shipley J.,		
RA	Cooper C.S.;		
RT	"The t(X;1)(p11.2;q21.2) translocation in papillary renal cell		
RT	carcinoma fuses a novel gene PRC3 to the TFE3 transcription factor		
RT	gene.",		
RL	Hum. Mol. Genet. 5:1333-1338(1996).		
RN	[3]		
RP	REVISIONS.		
RA	Clark J.;		
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	SEQUENCE OF 208-743 FROM N.A.		
RX	MEDLINE: 90249724.		
RA	Beckmann H., Su L.-K., Kadesch T.;		
RT	"TEF3: a helix-loop-helix protein that activates transcription		
RT	through the immunoglobulin enhancer muez motif.",		
RL	Genes Dev. 4:167-179(1990).		
RN	[5]		
RP	SEQUENCE OF 266-353 FROM N.A., AND CHROMOSOMAL TRANSLOCATION.		
RX	MEDLINE: 98054131.		
RA	Clark J., Lu Y.-J., Sidhar S.K., Parker C., Gill S., Smedley D.,		
RT	Hamoudi R., Linehan W.M., Shipley J., Cooper C.S.;		
RT	"Fusion of splicing factor genes PSF and Nono (p54nbp) to the TFE3		
RT	gene in papillary renal cell carcinoma.",		
RL	Oncogene 15:2233-2239(1997).		
CC	-1- FUNCTION: POSITIVE-ACTING TRANSCRIPTION FACTOR THAT BINDS TO THE		
CC	IMMUNOGLOBULIN ENHANCER MUEZ MOTIF. IT BINDS ALSO VERY WELL TO A		
CC	CC/MUTE SITE. BINDING OF TFE3 TO DNA INDUCES DNA BINDING.		
CC	-1- SUBUNIT: EFFICIENT DNA BINDING. REQUIRES DIMERIZATION WITH ANOTHER		
CC	BHLH PROTEIN.		
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.		
CC	-1- TISSUE SPECIFICITY: UBIQUITOUS IN FETAL AND ADULT TISSUES.		
CC	-1- DISEASE: INVOLVED IN PAPILLARY RENAL CELL CARCINOMA (PRCC) BY		
CC	CHROMOSOMAL TRANSLOCATIONS T(X;1)(P11.2;Q21.2) WHICH INVOLVES TFE3		
CC	AND PRC3: T(X;1)(P11.2;P34) WHICH INVOLVES TFE3 AND PSF, AND		
CC	INV(X)(P11.2;O12) THAT INVOLVES TFE3 AND NONO.		

```
CC -|- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X99721; CAA68061.1; -
DR EMBL; X97160; CAA65800.1; -
DR EMBL; X97161; CAA65800.1; JOINED.
DR EMBL; X97162; CAA65800.1; JOINED.
DR EMBL; X96717; CAA65478.1; -
DR EMBL; X51330; CAA35714.1; -
DR PIR; A34596; A34596.
DR PIR; S10379; S10379.
DR HSSP; P22415; IAN4.
DR TRANSFAC; T00811; -.
DR MIM; 314310; -.
DR PFAM; PF00010; HLH; 1.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW Chromosomal translocation; Proto-oncogene.
FT FT DOMAIN 260 271
FT STRONG TRANSCRIPTION ACTIVATION
FT (POTENTIAL).
FT BASIC DOMAIN.
FT DNA_BIND 344 359
FT DOMAIN 360 400
FT DOMAIN 409 430
FT DOMAIN 430 430
FT DOMAIN 575 743
FT SITE 178 179
FT BREAKPOINT FOR TRANSLLOCATION TO FORM
FT PRCC-TFE3 ONCOGENE.
FT SITE 295 296
FT BREAKPOINT FOR TRANSLLOCATION TO FORM
FT PSF-TFE3 ONCOGENE.
FT BREAKPOINT FOR TRANSLLOCATION TO FORM
FT NONO-TFE3 ONCOGENE.
FT GPKL -> EGRG (IN REF. 4).
FT P -> S (IN REF. 1).
FT P -> K (IN REF. 4).
FT P -> G (IN REF. 4).
FT T -> A (IN REF. 3).
FT A -> R (IN REF. 4).
FT V -> C (IN REF. 4).
FT MISSING (IN REF. 3).
FT ESSNGGPSGGLSASPS -> PAVSKASSRRSSFSWEES
FT CONFLICT 726 743
FT SEQUENCE 743 AA; 80007 MW; 1D82B94553543949 CRC64;
SQ
Query Match 1.8%; Score 143; DB 1; Length 743;
Best Local Similarity 33.8%; Pred. No. 1,82e+05;
Matches 26; Conservative 23; Mismatches 23; Indels 5; Gaps 4;
Db 346 OKKDHNLIERRRNRINDRIKELCTGLIPKSDEPMNKGITLKASVDYIRKLQKEQR 405
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 355 ERRTRHNLEIKKYRSINDRILOQLKVL--CGD-EAKSKSATLRALIEHFVENVHQV 411
Db 406 SKD-LDSRSLDEQANR 421
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 412 LKHHEVGMRKTL-QNNR 427
RESULT 9 STANDARD: PRT: 309 AA.
ID NO75_SOYBN P08297;
DT 01-JAN-1988 (Rel. 08, Created)
DT 01-AUG-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE EARLY MODULIN 75 PRECURSOR (N=75) (NGM-75).
EN ENOD2A AND ENOD2B.
GN Glycyne max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
```



```
CC TRANSCRIPTION FACTORS. "MYOGENIC FACTORS" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M69052; AAA30009.1; ALT_INIT.
DR PIR: A41123; A41123.
DR HSSP: P10085; IMDY.
DR TRANSFAC: T00926; -.
DR PFAM: PF01586; Basic; 1.
DR PFAM: PF00010; HLH; 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DNA_BIND 112 124 BASIC DOMAIN.
FT DOMAIN 125 164 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
SQ SEQUENCE 260 AA; 28682 MW; 8CC5FA9D2E6DAFA4 CRC64;

Query Match 1.6%; Score 128; DB 1; Length 260;
Best Local Similarity 33.8%; Pred. No. 2,96e-03;
Matches 23; Conservative 18; Mismatches 23; Indels 4; Gaps 4;

Db 112 DKRKATLRERRLR-KYNEAFELKRRHCANPQRLPKVEILRNAYIEKLERLQVE 170
QY 355 ERTNHNLEKK-YRCSINDRIQQLKVLGD-EAKLSKATLRATRIHIEVEHENQVL 412
Db 171 KANGSEM 178
QY 413 KHAVE-QM 419

RESULT 14
ID USF1_HUMAN STANDARD; PRT; 310 AA.
AC P22415;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE UPSTREAM STIMULATORY FACTOR 1 (MAJOR LATE TRANSCRIPTION FACTOR 1).
GN USF1 OR USF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-25.
RX Gregor P.D., Savadogo M., Roeder R.G.;
RA "The adenovirus major late transcription factor USF is a member of
RT the helix-loop-helix group of regulatory proteins and binds to DNA as
RT a dimer."
RL Genes Dev. 4:1730-1740(1990).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 197-260.
RX MEDLINE; 94139661.
RA Ferre-D'Amaré A.R., Pogononec P., Roeder R.G., Burley S.K.;
RT "Structure and function of the b/HLH/Z domain of USF."
RL EMO J. 13:180-189(1994).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO A SYMMETRICAL DNA
CC SEQUENCE (E-BOXES) (5'-CACGTG-3') THAT IS FOUND IN A VARIETY OF
CC VIRAL AND CELLULAR PROMOTERS.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. BINDS DNA AS AN HOMODIMER OR A HETERODIMER
CC (USF1/USF2).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55666; CA39201.1; -.
DR PIR: S13525; S13525.
DR PDB: 1AN4; 17-SEP-97.
DR TRANSFAC: T00874; -.
DR MIM: 191523; -.
DR PFAM: PF00010; HLH; 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
KW DNA-binding; Nuclear protein; Transcription regulation; 3D-structure.
FT DNA_BIND 200 212 BASIC DOMAIN.
FT DOMAIN 213 255 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 271 292 LEUCINE-ZIPPER (POTENTIAL).
SQ SEQUENCE 310 AA; 33538 MW; BFDA91519BA8B80AE CRC64;

Query Match 1.6%; Score 126; DB 1; Length 310;
Best Local Similarity 31.4%; Pred. No. 5,68e-03;
Matches 27; Conservative 25; Mismatches 27; Indels 7; Gaps 6;

Db 193 RTTDEKRRAGHNEVERRRRKINNIWVLSKIIPDCSMESTSGSGKGLSKACDYI 252
QY 350 RRPYTE-RRTHNLEKKRYRCSINDRIQQL-KVLL-CG-DEAK-LSKSATLRATRIHIE 403
Db 253 ELROSNNRLESELOGLDOLQDLDNDVL 278
QY 404 EVEHENQVLKHVEQMR-TLQNNRL 428

RESULT 15
ID USF1_RABIT STANDARD; PRT; 310 AA.
AC 002818; 002819;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE UPSTREAM STIMULATORY FACTOR 1 (MAJOR LATE TRANSCRIPTION FACTOR 1).
GN USF1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-NEW ZEALAND WHITE; TISSUE=LUNG;
RX MEDLINE; 97435316.
RA Gao E., Wang Y., Alcorn J.L., Mendelson C.R.;
RT "The basic helix-loop-helix-zipper transcription factor USF1
RT regulates expression of the surfactant protein-A gene."
RL J. Biol. Chem. 272:23398-23406(1997).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO A SYMMETRICAL DNA
CC SEQUENCE (E-BOXES) (5'-CACGTG-3') THAT IS FOUND IN A VARIETY OF
CC VIRAL AND CELLULAR PROMOTERS. REGULATES THE EXPRESSION OF THE
CC SURFACTANT PROTEIN-A (SP-A) GENE.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. BINDS DNA AS AN HOMODIMER OR A HETERODIMER
CC (USF1/USF2).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS: USF1A (SHOWN HERE) AND USF1B;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF003894; AAC48764.1; -.
DR EMBL: AF003895; AAC48765.1; -.
DR
```


THIS PAGE BLANK (USPTO)

RH 426 6466

RP 501 17